BIOMETRIKA

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

VOLUME XXXII

1941—1942

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

CONTENTS OF VOLUME XXXII

Memoirs

Miscellanea.

Vol. XXXII. Part I January, 1941

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

Reprinted by offset-litho, 1952, 1960

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[*Issued* 31 *January* 1941]

A THEORY OF RANDOMNESS

BY M. G. KENDALL

INTRODUCTION

1. In two recent papers Babington Smith & I (1938 and 1939) have discussed the problems of sampling with random numbers and the construction of tables of such numbers by mechanical methods. With the publication of 100,000 numbers (1940) what one may call the practical side of the investigation has come to an end. The purpose of this paper is to develop the theory of the subject and to put in their proper setting some of the ideas on which the practical research was based. It is divided into two parts. In the first I deal with the symbols and mathematics of the theory of random suites, my fundamental contention being that a theory of randomness can be developed within the framework of existing mathematical notions. The second part indicates how the theory is to be related to practice.

2. Much of the following work was suggested by the treatment of von Mises (1936) and Dörge (1934), and I take the opportunity of expressing my indebtedness to them. The principal difference between von Mises's views and my own concerns his concept of the Irregular Kollektiv, or infinite random series. Numerous attempts have been made to show that this concept leads to a contradiction and that it is therefore an improper foundation for a theory of probability. Such attempts have mostly failed, but under pressure of the criticisms embodied in them the definition of the Irregular Kollektiv has been successively modified by von Mises's followers until it has lost the pristine simplicity which was originally one of its most attractive features. I do not propose to discuss here the difficulties associated with the concept of the Irregular Kollektiv or the various expedients which have been proposed to meet them. I have tried to cut the Gordian knot by rejecting the concept, and the theory below accordingly avoids all the difficulties attendant upon it.

PART 1. THE THEORY OF RANDOM SUITES

3. I consider a finite number r of symbols $A_1, A_2, \ldots, A_r$, each of which will be called a characteristic, and an infinite ordered series of these characteristics, which will be called a *suite*. For instance, if there were two characteristics A_1 and A_2 such a suite might be

$$A_1 A_2 A_1 A_2 A_1 A_2 A_1 A_2 \ldots, \qquad \ldots\ldots(1)$$

where the characteristics appear alternately. Suites exist in the sense that they

can be completely specified by a law of formation, as the foregoing example shows.

4. *Definition.* If the proportional frequency of each characteristic in a suite tends to a limit in the mathematical sense, the suite is called "proper". In the contrary case, "improper".

Proper suites exist; e.g. the proportional frequencies of A_1's and A_2's in (1) tend each to the limit $\frac{1}{2}$.

Improper suites also exist. For example, it may be shown that if we take the kth digit in the logarithms to base 10 of all the integers, beginning with 1, the suite of the digits 0–9 so obtained is improper, since no proportional frequency tends to a limit.

Suites also exist which are proper for one characteristic and improper for another, provided that there are more than two characteristics. For we may build a suite from the logarithm table in the manner just described, and then insert a new characteristic Q between successive digits. The proportional frequency of Q will then tend to $\frac{1}{2}$, but those of the others will not tend to a limit. If, however, there are two characteristics, the proportional frequencies, being together equal to unity, must tend to a limit together.

5. *Definition.* The limit of the proportional frequency of a characteristic in a proper suite is called the probability of that characteristic in that suite.

6. *Definition.* By a "Selector" I mean an infinite series of positive integers ordered according to their magnitude. A selector, being infinite, must be specified by a law of formation, not by enumeration.

There is a special class of such laws which deserves separate consideration. Suppose we have such a suite as this:

$$A_1 A_2 A_2 A_1 A_1 A_2 A_2 A_1 A_1 A_2 A_2 A_1 \ldots, \qquad \ldots\ldots(2)$$

characteristics after the first appearing alternately in pairs.

Our law of formation of the selector might be in these terms: proceed along the series until you come to the combination $A_2 A_1 A_1 A_2$; then choose the ordinal number of the next following member of the series, and proceed until you again meet that combination; and so on. The series of ordinals so obtained is the selector.

The importance of selectors of this type is that they are mathematically independent of the particular characteristic of the member whose ordinal number is chosen. By mathematically independent I mean that the value of this member does not appear in the law of formation, so that the same member would be chosen whatever its characteristic.

Definition. If a selector is constructed from a suite and, in virtue of the law of formation, any member of the selector is mathematically independent of the

characteristic whose ordinal number in the suite is the value of that member, the selector is said to be "disjoint" with respect to the suite.

7. It might happen that a law of formation of a disjoint selector was given which did not in fact lead to a selector in the case of certain suites. For example, with the suite (2), if we try to construct a selector by choosing ordinals corresponding to the characteristics following three successive A_1's, no ordinals appear. Such a law I should regard as degenerate in relation to that suite, and I exclude it from the domain of discussion from this point onwards. Hereafter, in speaking of a selector in relation to a given suite I shall assume that the one is disjoint with respect to the other.

8. We may now apply selectors to pick out subsets from a suite. We do so by choosing from the suite those members whose ordinals are the numbers appearing in the selector.

Ex hypothesi, the result of this process will be a new suite of the characteristics (some at least) of the original suite. We may call this a "Derived suite". Symbolically, denoting the selector by the roman S and the suite by K, we may write

$$D = \mathrm{S}K. \qquad \qquad \text{......(3)}$$

I proceed to prove one or two theorems of a negative kind about derived suites.

9. A suite derived from a proper suite is not necessarily proper.

For let $K = A_1 A_2 A_1 A_2 A_1 A_2 A_1 A_2 \ldots,$

$\mathrm{S} = 1, 2, 4, 5, 7, 9, 11, 13, 15, 16, 18, 20, 22, \ldots,$

the numbers running alternately in sets of even and odd, the number of each kind being equal to twice the number of preceding members of the selector. Then

$$\mathrm{S}K = A_1 A_2 A_2 A_1 A_1 A_1 A_1 A_1 A_1 A_2 A_2 A_2 A_2 A_2 A_2 \ldots.$$

However far we go in this series, say to the end of a run of A_1's, there will follow twice as many A_2's as there have already occurred of both A_1's and A_2's. Clearly the suite is improper.

10. If we apply another selector S_2 to a derived suite we get a further derived suite which we may write $\mathrm{S}_2\mathrm{S}_1K$. It is clear that this will not in general be the same as $\mathrm{S}_1\mathrm{S}_2K$.

The "identical" selector $\mathrm{E} = 1, 2, 3, 4, \ldots$ is of some importance. Clearly it reproduces a suite to which it is applied, and $\mathrm{ES}_1K = \mathrm{S}_1\mathrm{E}K$, etc.

Randomness

11. *Definition.* If the probability of the characteristic A_1 in a proper suite is p; and if the probability of A_1 in a proper suite derived from it by the selector S is also p; then the suite is said to be *random* for the characteristic A_1 with respect to S.

Suites and selectors with this property exist. Every proper suite is random

with respect to the identical selector E, and the suite (1) is random for both A_1 and A_2 with respect to the selector

$$1, 4, 7, 10, \ldots,$$

though not to the selector $\quad 1, 3, 5, 7, \ldots.$

12. *Definition.* A suite which is random for a characteristic A with respect to a number of selectors $S_1, S_2, \ldots, S_m$ is said to be random in the selector domain $S_1, S_2, \ldots, S_m$.

It is to be noted that if a suite is random with respect to S_1 and S_2 it does not follow that S_1K is random with respect to S_2 or S_2K with respect to S_1. E.g. if

$K = A_1A_1A_2A_2A_1A_1A_2A_2 \ldots,$

$S_1 = 1, 3, 5, 7, 9, \ldots,$

$S_2 =$ the disjoint selector obtained by writing down the ordinals of characteristics next following A_1,

then
$$S_1K = A_1A_2A_1A_2A_1A_2 \ldots,$$
$$S_2K = A_1A_2A_1A_2A_1A_2 \ldots,$$

so that K is random for A_1 and A_2 with respect to both S_1 and S_2. But

$$S_2S_1K = A_2A_2A_2A_2A_2A_2 \ldots,$$
$$S_1S_2K = A_1A_1A_1A_1A_1A_1 \ldots.$$

13. Given a suite and a certain finite set of selectors, we may consider the suites obtained by repeated applications of groups of these selectors. This will give us a series of derived suites which may be infinite but is nevertheless ordered. If all the resulting suites are proper and the probability of a characteristic A in them all is the same as that in the parent suite, the latter is said to be *completely* random in the selector domain $S_1, S_2, \ldots, S_m$.

There exist suites which are completely random in certain domains. E.g. if

$$K = A_1A_2A_1A_2A_1A_2 \ldots,$$
$$S_1 = 1, 4, 7, 10, \ldots,$$

then
$$S_1K = A_1A_2A_1A_2A_1A_2 \ldots = K,$$

so that repeated applications of S_1 lead only back to the original suite.

Consider further $\quad S_2 = 2, 5, 8, 11, \ldots.$

Then
$$S_2K = A_2A_1A_2A_1A_2A_1 \ldots$$

and
$$S_2^2K = A_1A_2A_1A_2A_1A_2 \ldots.$$

Thus, any number of applications of S_1 and S_2 lead either to $A_1A_2A_1A_2A_1A_2\ldots$ or to $A_2A_1A_2A_1A_2A_1\ldots$, and hence the suite is completely random for A_1 and A_2 with respect to S_1 and S_2.

It follows at once that any suite which is derived from a completely random suite by a selector of the set is also completely random within the same domain.

14. The foregoing examples, though trivial enough, show that the various ideas which have been introduced are not self-contradictory, and that they fall within the scope of ordinary mathematical concepts. But the use of the word "random" to describe a state of affairs which is the reverse of what is ordinarily understood by the word requires some scholium. I have, for instance, remarked that any suite is random with respect to the identical selector 1, 2, 3, 4, ... But surely, it may be said, such a suite as $A_1A_1A_1A_1A_1A_1A_1A_1A_1\ldots$ is the very reverse of random, being as systematic as any such series can be? I will anticipate a later part of this paper to some extent by a short explanation of this point.

15. In statistical work we require one thing above all in a "random" selection; namely, that if continued long enough it shall draw all members of the universe equally often, or at least in a known proportion. In fact, it is not the haphazard quality of randomness that we use in drawing inferences from random samples, but the only thing about it which is not haphazard, namely its property of producing definite limits. (I am, of course, speaking colloquially.) Any method of selection would serve equally well if it satisfied this primary requisite. The "random" series with which we are familiar in ordinary work are unpurposive and chaotic in appearance for two reasons: firstly, because we fondly hope to have a series which gives a suite random in regard to all possible selectors, so that it can be used to draw random samples from all universes whatever the characteristic under consideration; such a series must be random in a very wide selector domain, including all the more obvious selectors which would give the series a purposive appearance, and consequently it looks unsystematic; secondly, as an experimental fact we have learnt that when a sample is chosen haphazardly it is often random, at least so nearly so that ordinary inspection of a series of results will not reveal the difference; this haphazard selection leads us to expect from it an unpurposive-looking result.

16. But there is no virtue in lack of purpose for its own sake. In fact, random sampling has a very definite purpose, and it is the purposive parts of it that we have in mind in using the method at all. As the domain of selectors becomes larger, the random suite becomes more and more like the completely haphazard entity which von Mises would like to make the basis of his theory; but in my view the random suite must always be considered as random in a finite domain. I contend that there is no such thing as absolute randomness, just as there is no such thing as absolute velocity. The latter has meaning only with reference to a co-ordinate framework, the former only with reference to a selector framework. I might summarize the attitude of the foregoing paragraphs by saying that they are founded on the concept of the *relativity* of randomness. If this be agreed, the difficulties about terming "random" certain series which do not conform to the colloquial use of the word at once disappear.

Multi-dimensional suites

17. As a simple extension of the idea of a suite of characteristics we may consider suites of sets of characteristics. Such an extension offers no difficulty, and is very similar to the transition from describing points on a line in terms of one co-ordinate to points in a multi-dimensional space by several co-ordinates.

We may amalgamate two or more suites into a suite of more dimensions. E.g. with the suites

$$A_1 A_2 A_1 A_2 A_1 A_2 A_1 A_2 \ldots,$$
$$B_1 B_2 B_3 B_1 B_2 B_3 B_1 B_2 B_3 \ldots,$$

we can associate the nth member of one with the nth member of the other to obtain

$$(A_1 B_1)\,(A_2 B_2)\,(A_1 B_3)\,(A_2 B_1)\,(A_1 B_2)\,(A_2 B_3)\,\ldots.$$

Convolution

18. We may also construct an m-dimensional suite by dividing a one-dimensional suite into blocks of m. This process is worth noticing. Consider the suite

$$A_1 A_2 A_2 A_1 A_1 A_2 A_2 A_1 A_1 \ldots.$$

This is proper and each characteristic is random with respect to the selector

$$1, 3, 5, 7, 9, \ldots.$$

Now suppose we make a two-dimensional suite by bracketing successive terms, thus:

$$(A_1 A_2)\,(A_2 A_1)\,(A_1 A_2).(A_2 A_1)\ldots.$$

This is proper with respect to the two two-dimensional characteristics $(A_1 A_2)$ and $(A_2 A_1)$ but it is not random with respect to the selector given.

Definition. I shall refer to the process of deriving a multi-dimensional suite from a one-dimensional suite by grouping sets of successive terms as "convolution", and the derived suite will be said to be "convoluted". From the example given it is clear that a convoluted suite is not necessarily random in the domain of randomness of the parent suite.

Independence

19. *Definition.* If a two-dimensional suite is derived from two one-dimensional suites by attaching one member of the first to the member with the same ordinal number in the second; if the original suites and the new suite are proper; and if the probability in the derived suite of a characteristic $(A_j B_k)$ is the product of the probabilities of A_j in the first suite and of B_k in the second for all j and k; then the two original suites are said to be statistically independent.

Definition. If from a proper two-dimensional suite there are derived two proper one-dimensional suites by ignoring the first and then the second characteristic of the pairs which constitute the suite; and if these two suites are statistically independent, the two sets of characteristics are said to be statistically independent in the original suite.

Statistical independence as thus defined concerns either suites or characteristics in suites. Like probability it is a property of aggregates, not of individuals.

20. The generalizations of statistical independence to the case of several suites or multi-dimensional suites can be made without difficulty. I shall here omit them and the theorems which they obey, since all the results are obvious extensions of the theory of class frequencies set out for the case of finite classes in the *Introduction* by Udny Yule and myself (1939). The following results are, however, worth recalling:

(*a*) If K, L, M are proper suites, K is statistically independent of L and L is statistically independent of M; it does not follow that K is statistically independent of M.

(*b*) Three suites are statistically independent only if the probabilities $(A_j B_k C_l)$ are equal to the product of the probabilities A_j in K, B_k in L, and C_l in M. It is not sufficient that they should be statistically independent pair and pair, as the following example shows:

$$K = A_1 A_2 A_1 A_2 A_1 A_2 A_1 A_2 \ldots,$$
$$L = B_1 B_2 B_2 B_1 B_1 B_2 B_2 B_1 \ldots,$$
$$M = C_1 C_1 C_2 C_2 C_1 C_1 C_2 C_2 \ldots.$$

Here, for instance, the probability of $(A_1 B_1 C_2)$ is zero in the suite obtained by associating triads from members of the suites which have the same ordinal.

Local randomness

21. A suite as defined above is infinite. I now consider a finite series of characteristics which I call a *sequence*. A sequence may be considered as a section of a suite.

It is evident that any sequence, being finite, can form part of a suite in which the probabilities have any given value and which is random in any assigned domain. We may, however, imagine the selectors of the domain applied to the sequence, that part of the selectors which contains numbers greater than the number of members in the sequence being ignored. Similarly, we can convolute the sequence in any way consistent with its size and apply selectors to the sequences so derived. We can compare the actual proportions of characteristics in these sequences with those in any given suite. Any such process I call a test.

Definition. If the proportional frequencies in a sequence are approximately what they would be in a suite random with respect to the selectors of the test, the sequence is said to be locally random with respect to that test; and so for a test domain.

To make this definition precise it is necessary to consider what is meant by "approximately". Suppose the sequence is of size n, and consider the r^n possible

sequences of this size. To each there will correspond a proportional frequency under the tests. Choose a number of these, αr^n, which may be regarded as "approximately" the same as the proportional frequency in the suite. Then if the given sequence is one of these, it is "approximately" the same. Clearly the word approximately depends on the choice of the number α which corresponds to what is generally known as a "level of significance".

The concept of local randomness, in my view, is important. The series of characteristics which we encounter in real life are always sequences, not suites; and we have to estimate probabilities and random properties from finite aggregates, not from infinite series such as form the basis of the theory. Any finite series of characteristics whatever is random in the sense that it *might* arise, however infrequently, in random sampling. But in order to make any practical use of our theory we have to consider certain series as non-random, or in other words we have to judge from the local randomness of observed sequences.

Part 2. Application of the theory to practice

Events

22. Events are the primary data of statistical experience. Every event has a number of properties, the conceptual abstractions of the Gestalts which it provides. These properties may be called characteristics, and it is with aggregates of characteristics that statistical inference is concerned. The throwing of a die and the growing of a crop on a given field are events. Characteristics of the former would include the number which came uppermost, the time at which the throw was made, the angles which the edges made with a line fixed in space, and so on. In general an event has an infinite number of characteristics. When we have a complex phenomenon such as a crop of wheat on a field it is a matter of choice whether we regard the whole thing as one event, or look on it as a series of associated events, e.g. a collection of crops on a number of square yards. But the event is to be regarded as including the whole of the happening, and is not synonymous with one of its characteristics. A yield of wheat is not an event, nor is the number thrown by a die.

23. Consider then an aggregate of events. Suppose there exists a finite set of characteristics such that each event has one and only one characteristic of the set; for example, the events consisting of throws of an ordinary die must have one of the characteristics 1–6, according to the number which falls uppermost, and cannot have more than one. We can then say that the aggregate of events gives rise to an aggregate of characteristics.

24. Now the aggregates of events we meet in experience are always finite. It is true that we sometimes regard a line as composed of an infinite number of points or a solid body as an infinite number of particles; but these are mental fictions and it is not possible to observe the characteristics of an infinite number

of events. The finite aggregates of our experience can be ordered to produce a sequence—in fact they are usually arranged for us by the temporal order in which they occur. The fundamental problem of linking theory and practice—and an analogous problem arises in all frequency theories of probability—is to relate the sequences of observation with the suites of theory.

25. The sequences of observation may be regarded as generated by a physical process. The sequence consisting of throws of a die, for example, may be considered as defined by the rules under which the die is cast. A sequence of crop yields is determined by the circumstances under which the crop was grown. I shall assume that the physical process generating a sequence can be depicted by a mathematical law defining a suite, in the same way that the "straight lines" we draw on paper can be depicted by the straight lines of Euclidean geometry, or a rigid body by the abstractions of mathematical dynamics. Members of a sequence are ascertained by experiment; those of a suite by calculation.

26. I also take it as empirically established that there are observational sequences which can be adequately described as locally random sections of suites; random, that is, in certain domains. And I assume that the processes generating these sequences will, if continued, produce further sequences which are also locally random. This is essential to all scientific inquiry, that a law which is established will continue to operate. If it does not, we must alter the law; but before carrying out the extra trials we can only act on the assumption that the law will hold. Put in this way, perhaps, the assumption seems unjustified, but it is made every moment of our lives. In writing these words I am assuming that a past phenomenon will recur, namely that a particular arrangement of marks on a piece of paper will evoke certain ideas in the reader's mind. This much I am compelled to concede to those writers, like Mr Keynes and Dr Jeffreys, who contend that probability cannot be defined in terms of frequency, namely that the uncertain attitude of mind which one adopts towards some laws cannot be measured by probability as I have defined it. As to the phenomenon of mental doubt, the scientific procedure by which it is removed or strengthened, the desirability of measuring it, I am in agreement with Dr Jeffreys; but I do not call the measure probability in a statistical context. This is only a matter of words, but unfortunately so is a great deal of statistical discussion.

27. The probability of a characteristic in a physical process is to be estimated from the observed sequence to which the process gives rise. I need not dwell here on the methods of estimation and the ideas which underlie them. But there is one point of some importance to note. It is evident that probability is a property of characteristics, not of events, and to be strictly accurate we should always speak of it as such. For instance, if I toss a penny on to a chessboard, the

limit of the proportion of heads may be $\frac{1}{2}$, but the limit of the proportion of cases in which it falls on a white square may be $\frac{3}{4}$. Neither of these fractions is the probability of the event. They are probabilities of characteristics and there is nothing inconsistent in the fact that they are different.

Independence

28. The statistical independence of two suites, or of characteristics in a multi-dimensional suite, was defined in paragraph 19, and the statistical independence of observed sequences follows the same line. In statistics the question whether two series of characteristics are independent is to be determined purely from the experimental data. There may be very good reasons why one characteristic is "dependent" on another in a causal sense, but if the occurrence of one is not accompanied by the occurrence of the other in "unexpected" proportion they are statistically independent. Contrariwise, there may be no obvious causal nexus and yet the two may be statistically dependent. In fact, I would be inclined to deny any separate meaning to "causal" dependence other than that of statistical dependence (with perhaps, allowance for the temporal element).

29. This point is important in one respect. I have up to the present spoken only of the independence of characteristics, not of events, and even of the former only in terms of suites or sequences. But in the theory of probability as expounded in textbooks it is quite common to meet with such expressions as "a series of independent events", or "successive events are independent". The word "event" here means what I call a characteristic; but can we speak of "a suite of independent characteristics"? I do not think so. In my opinion the concept is equivalent to that of the Irregular Kollektiv of von Mises.

30. For example, one would be inclined to begin an approach to a definition of the concept by requiring that each characteristic was followed equally frequently by all characteristics of the suite, e.g. that in a suite of characteristics $A_1 A_2$, an A_1 was followed equally frequently by an A_1 or an A_2. But this is true of the suite

$$A_1 A_1 A_2 A_2 A_1 A_1 A_2 A_2 \ldots,$$

which is clearly not of the type desired. One might then require that each characteristic should, in addition, be followed next but one by all other characteristics in equal amount. But this is true of the suite consisting of repetitions of

$$A_1 A_2 A_2 A_2 A_1 A_1 A_1 A_2 \ldots.$$

Baffled by continual examples of this kind, one might then require that the occurrence of any characteristic was to be independent of all or any of the characteristics which have preceded it. This, on analysis, is found to be equivalent to the requirement that the suite shall be random with respect to all disjoint selectors of the type considered in paragraph 6; and this is precisely the difficulty of the Irregular Kollektiv.

31. I can see no way round this difficulty; and I therefore reject the suite of independent events as I reject the Irregular Kollektiv. This has two important consequences, the first concerning the ordinary theorems of probability and the second concerning Bernoulli's theorem.

The first point may best be illustrated by an example: suppose the probability of getting a head with a toss of a penny is $\frac{1}{2}$. What is the probability of getting two heads with consecutive tosses? Anyone grounded in the classical theory would answer "$\frac{1}{4}$" without hesitation. Nevertheless, the result is only true under certain conditions. In fact the data of the problem are that there is a suite of throws of the penny and that the proportional frequency of heads in this suite is $\frac{1}{2}$. Now such a suite might be (A_1 = heads, A_2 = tails)

$$A_1 A_2 A_1 A_2 A_1 A_2 \ldots,$$

and the probability of getting two successive heads is zero. But, it may be objected, this is an artificial series which would never occur. To this I should reply, agreed, but why should there not occur a natural series in which the proportional frequency of pairs of heads did not tend to $\frac{1}{4}$? It will, I hope, be clear on reflection that there is nothing in the data of the problem to require the answer $\frac{1}{4}$ as a logical necessity unless we make some additional assumption such as this: the occurrence of one characteristic is statistically independent of the occurrence of the next. This contains the answer of $\frac{1}{4}$ implicitly.

32. In generalization of this problem we might ask: if the probability of the characteristic is p, what is the probability that in a set of n characteristics we shall get r successes and $n-r$ failures? Here, again, the answer of the classical theory would be $\binom{n}{r} p^r (1-p)^{n-r}$; and here again the result is only true if we assume the statistical independence of sets of n. Clearly if the result is to be true for all n we are once more verging on the suite of independent characteristics referred to in paragraph 30. I conclude that for statistical purposes the results of the classical theory of probability are not to be accepted without examination. If in any particular case we require one of these results, we must be satisfied that the suite we are considering is such as to justify the use of it.

Bernoulli's theorem

33. The well-known theorem given by James Bernoulli in the *Ars Coniectandi* is subject to similar limitations in regard to its statistical applications. In essence the theorem is a proposition in algebra which may be stated thus: in the binomial distribution $(p+q)^n$, if u be the sum of the greatest term and the n preceding and n succeeding terms, the ratio of u to the sum of the remaining terms may be made as large as we please by increasing n sufficiently. There can be no criticism of this result. But, as applied to statistical series, the theorem

states that if the probability of a characteristic is p and we observe m sets of n events, the proportion of sets in which the proportion of successes differs from p by less than ϵ tends to unity with large m and large n. Or put another way, the probability that the proportion of successes in a set of m differs from p by less than ϵ tends to unity as m tends to infinity. Symbolically,

$$P\{|\,p-h(A)\,|<\epsilon\}>1-\eta, \qquad m>M,$$

where $h(A)$ is the proportion of successes.

34. This is a statement about the probability of a probability and I need not emphasize the logical weakness inherent in it. On looking into the proposition further we find that it is dependent on the type of assumption considered in paragraph 32, namely, that if the probability of a characteristic is p, the probability of r characteristics in a set of n is $\binom{n}{r} p^r (1-p)^{n-r}$.

This will only be true if the observed series is, approximately at least, a series of "independent" characteristics. Consider, for example, the suite

$$A_1 A_1 A_2 A_2 A_1 A_1 A_2 A_2 \ldots,$$

the characteristics of which are random with respect to the selector

$$1, 3, 5, 7, 9, \ldots$$

and have probabilities each equal to $\frac{1}{2}$. Consider the convoluted suite

$$(A_1 A_1)\,(A_2 A_2)\,(A_1 A_1)\,(A_2 A_2).$$

Bernoulli's theorem, as usually stated, would lead us to the conclusion that the probability of getting in this suite a pair containing one A_1 and one A_2 is $\frac{1}{2}$. Actually it is zero.

If, once again, it is objected that this is a highly artificial series, I reply as before that series with the same properties might arise naturally. We can only make Bernoulli's theorem legitimate by postulating that the suite to which it is applied shall have the property of randomness under convolution for all n. I do not think this is always a legitimate hypothesis to make, but I am anxious not to be misunderstood on the point. There undoubtedly exist sequences which can be regarded as belonging to suites random in a very wide domain—so wide that for many practical purposes they can be taken to be sequences of "independent" events. The point to be stressed is that this assumption underlies a great deal of statistical work but is never brought to light, and indeed is often not realized. The statistician takes it for granted; but to the philosopher nothing is more surprising than the orderly disorder which is common in Nature.

Random sampling

35. A statistical universe is an aggregate of objects, which may be finite or infinite. I consider selective processes which consist of abstracting one member at a time from this universe, and I suppose that each member is returned to the

universe after drawing if the universe is finite. The abstraction of each member may then be considered as an event, whose characteristics may be noted. I assume that there exists for these events a set of characteristics such that each event must bear one characteristic and can bear only one.

We may then imagine this selective process, which I shall call *sampling*, as a generator of a sequence of any desired extent, and to be capable of continuation without limit. The result of unterminating sampling would be to give a suite of events, to which there correspond one or more suites of characteristics. In practice we shall have a sequence only.

36. *Definition.* If a sequence obtained from a universe U by sampling is locally random for a characteristic A within a selector domain D, then the sampling is said to be random for U with respect to A within the domain D; and any member of the sequence is called a random sample for the characteristic A in the domain D.

This definition brings out the extremely relative nature of random sampling. A method which is random for one universe may not be random for another; a method random for one characteristic may not be random for another, even in the same universe; *and the randomness is always relative to the selector domain.* It is also to be noticed that the sampling process, being physical, can only be related to sequences, not to suites.

The assumption we make in using a random sampling method is that if it has in the past generated locally random sequences it will continue to do so in similar circumstances. The justification for this assumption is empirical.

37. In practice we sometimes draw samples one at a time and so obtain a one-dimensional sequence. We then convolute this sequence into groups of n, making an n-dimensional sequence. But we may also draw the samples in a block of n (which I shall call a "clutch"). The difference is of some importance. If we ignore the order of the individuals in a convoluted sequence we have what is virtually a clutch, and it is very common in statistical work to ignore the order in this way. A series of sampling results, for instance, are frequently given without any indication of the order in which the individual results appeared. It should not be overlooked that certain information relevant to the randomness of the sample has disappeared in the process. For example, we may be told that in a sample of 1000 births 510 were male. We should probably conclude that there was nothing in the sample to show that it was not random. But if we know that the first 510 were male we should certainly conclude that it was not.

38. What are the grounds on which a selective process of experience is considered to give random samples? In the first place, as has already been remarked, we can only use a selective process to produce a finite sequence. This sequence is always locally random in some domain or other. If we find that, as the sequence is increased, local randomness is maintained, we may say that the

method is random for the universe and characteristic considered. But we require more for a method to be used in practice. We require to be able to suppose with some confidence *a priori* that it will be random for fresh inquiries, fresh universes, fresh characteristics and fresh sequences. And we require the domain of randomness to be as wide as possible. It was formerly the custom to assume randomness to any desired extent if there was no obvious reason to the contrary—a sort of Principle of Non-sufficient Reason. This is most unsatisfactory, and could only be justified if it was found in practice that haphazard methods of selection give locally random sequences. In fact we find that whenever any element of personal choice is allowed free play, bias is very liable to appear.

39. It seems to me that we can never rid ourselves entirely of the possibility that a method of selection may lack randomness; but we can safeguard against the possibility to a great extent. For instance, the method of Random Sampling Numbers applied to a universe of names in a directory gives us something near certainty (if I may be allowed that colloquial expression) that the resulting samples will be random. Furthermore, we can experiment with a method to see if bias has appeared. If it has not, we are justified in expecting that it is random for the class of cases in which it has been tried. Ultimately, however, the assumption of randomness is part of the hypothesis which is being tested.

40. An assumption which is usually made in practice is that the method is random within whatever domain happens to suit the investigator at the moment. One draws a random sample from the universe of inhabitants of the British Isles. One says that the sample is "random" without any qualification. Behind this lies the assumption of the Irregular Kollektiv which has been considered from a different angle in paragraph 30. A great many statisticians would use such a sample to test any hypothesis about the universe which they chanced to encounter; they would assume that it was random in regard to height, sex, age or any other characteristic; they would assume that it was random under convolution; and they would assume the legitimacy of testing in any sampling distribution which happened to be convenient. All of this amounts to an assumption of randomness in a very wide domain, depending on a subjective judgment which may be quite wrong. The wider the domain, the less likely (again speaking colloquially) is the assumption to be justified. In practice this assumption frequently has to be made, and can be made without much danger with a good sampling method. But the greatest danger lies in the fact that the person making the assumption very often does not even realize that he is doing so. In any sampling inquiry it is necessary to ask oneself, Is the sampling method I am using random for the universe I am considering, for the characteristics I am discussing, and for the sampling distributions or tests of significance I am employing? Randomness is relative.

REFERENCES

DÖRGE, K. (1934). "Eine Axiomatisierung der von Miseschen Wahrscheinlichkeitstheorie." *Jber. dtsch. MatVer.* **43**, 39.

KENDALL, M. G. & BABINGTON SMITH, B. (1938). "Randomness and random sampling numbers." *J.R. Statist. Soc.* **101**, 147.

—— —— (1939). "Second paper on random sampling numbers." *Supp. J.R. Statist. Soc.* **6**, 51.

—— —— (1940). *Tables of Random Sampling Numbers.* Cambridge University Press.

VON MISES, R. (1936). *Wahrscheinlichkeit, Statistik und Wahrheit.* Springer, Berlin.

YULE, G. UDNY & KENDALL, M. G. (1939). *An Introduction to the Theory of Statistics.* 12th edition. Griffin & Co., London.

A GEOMETRICAL ANALYSIS OF THE FREQUENCY DISTRIBUTION OF THE RATIO BETWEEN TWO VARIABLES

By C. NICHOLSON, M.C., M.A., M.D.

This subject has already been treated by Geary (1930), and by Fieller (1932), the approach to the problem in both cases being algebraical; the geometrical approach to the same problem suggested itself to me when I was working on a series of anatomical measurements (Nicholson, 1938). This paper could hardly have reached publication without the generous assistance of Mr N. L. Johnson of University College, London.

(1) Variables independent

We are to consider the distribution of the ratio

$$\frac{y+\bar{Y}}{x+\bar{X}},$$

where $\bar{Y}$ and $\bar{X}$ are constants, and the joint distribution of x and y is given by the normal bivariate surface

$$z=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left\{\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right\}\right].$$

Then if we refer our observed values $(y+\bar{Y})$ and $(x+\bar{X})$ to the axes $y=0$ and $x=0$, the co-ordinates of the intersection of the zero values of the variables will be $-\bar{X}$ and $-\bar{Y}$, since $x+\bar{X}=0$ when $x=-\bar{X}$. The ordinates for a constant value of the ratio will lie in a plane surface passing through this point of the general form

$$y=mx+c,$$

where m will have the value of the ratio, and c will be $m\bar{X}-\bar{Y}$. The equation to the projection of the section of the normal surface by this plane on the (x, z) plane will be

$$z=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[\frac{1}{2}\left\{\frac{x^2}{\sigma_x^2}+\frac{(mx+c)^2}{\sigma_y^2}\right\}\right] \tag{1}$$

which can be rearranged to

$$z=\frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left\{\frac{c}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}\right\}^2\right]\exp\left[-\frac{1}{2}\left\{\frac{x+\dfrac{mc\sigma_x^2}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}}{\dfrac{\sigma_x\sigma_y}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}}\right\}^2\right] \tag{1a}$$

The area of this section is the integration of z from $-\infty$ to $+\infty$, the variable (because of the change of angle) being $x\sqrt{(1+m^2)}$, and this is

$$\frac{\sqrt{(1+m^2)}}{\sqrt{(2\pi)}\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}\exp\left[-\frac{1}{2}\left\{\frac{c}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}\right\}^2\right]. \tag{2}$$

The quantity $\dfrac{c}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}$ in (2) is equal to $\dfrac{m\bar{X}-\bar{Y}}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}}$, which is the function of the ratio which Geary treats as t. Now m is the tangent of an angle, say β, and (2) may be put in the form

$$\frac{1}{\sqrt{(2\pi)}\sqrt{(\sigma_x^2\sin^2\beta+\sigma_y^2\cos^2\beta)}}\exp\left[-\frac{1}{2}\left\{\frac{c\cos\beta}{\sqrt{(\sigma_x^2\sin^2\beta+\sigma_y^2\cos^2\beta)}}\right\}^2\right]. \tag{3}$$

Here $c\cos\beta$ is the perpendicular distance from the origin of the surface to the plane $y=mx+c$, so that we can draw the conclusion that a series of parallel plane sections of a normal surface making an angle of β with the x axis are normal curves with a standard deviation of

$$\frac{\sigma_x\sigma_y}{\sqrt{(\sigma_x^2\sin^2\beta+\sigma_y^2\cos^2\beta)}},$$

while their areas form another normal curve with the variable $c\cos\beta$ and with a standard deviation of

$$\sqrt{(\sigma_x^2\sin^2\beta+\sigma_y^2\cos^2\beta)}.$$

(2) VARIABLES CORRELATED

Clearly, in the case where the primary variables are not independent, it is still possible to reduce the distribution to this same geometrical system, so that we may discard reference to the primary variables and refer rather to the principal axes of the surface generalizing (3) as

$$\frac{1}{\sqrt{(2\pi)}\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}} \times\exp\left[-\frac{1}{2}\left\{\frac{k\sin\theta}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}\right\}^2\right], \tag{4}$$

where, referring to Fig. 1, a and b are the standard deviations in the direction of the major and minor axes respectively of a normal surface; also k is the distance of a point K from the origin, K being the focus of a pencil of planes cutting the surface, and k being equal to $\sqrt{(\bar{X}^2+\bar{Y}^2)}$. The angle KOA is α, which is the absolute value of the angle less than $\frac{1}{2}\pi$ between the major axis of the surface and the line joining the origin to the intersection of the zero value of the variables; it is

$$|\tan^{-1}(\bar{Y}/\bar{X})-\delta|,$$

where δ is the angle between the x axis and the major axis which is given by the equation

$$\tan 2\delta=\frac{2\rho_{xy}\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}.$$

θ is the angular deviation of any plane from the angle α which is taken as the origin of the pencil, and the value of any ratio will be given by $\tan(\alpha+\delta+\theta)$.

It is clear that if we were to confine ourselves to the distribution of the ratio we must use $\tan(\alpha+\delta+\theta)$ as the variable, but in this generalized form it is more informative to use the angle itself as the variable. The cumulative frequency for a deviate θ is the content of the two plane angles PKT and $P'KT'$, and where K lies without the bulk of the distribution the content of the angle PKT will be negligible. The content of the angle $P'KT'$ may then be taken as the content between

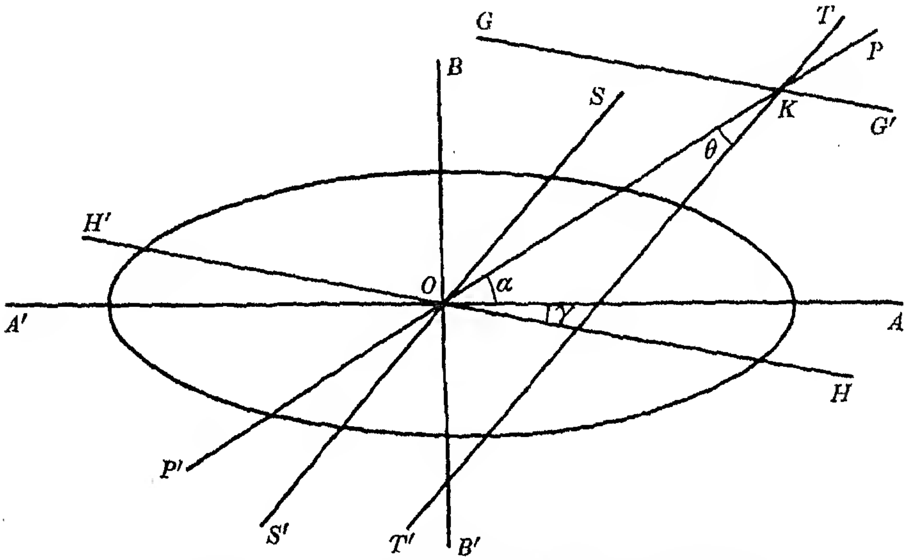


Fig. 1. Projection of normal bivariate surface to illustrate the geometry of the ratio between variables.

the two parallel planes, TKT' and SOS', the latter passing through the origin of the surface, and this is

$$\frac{1}{\sqrt{(2\pi)}\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}} \times\int_0^{k\sin\theta}\exp\left[-\frac{1}{2}\left\{\frac{u}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}\right\}^2\right]du. \quad (5)$$

If in (5) we put $\dfrac{u}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}$ as t, we get

$$\frac{1}{\sqrt{(2\pi)}}\int_0^{\frac{k\sin\theta}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}} e^{-\frac{1}{2}t^2}dt, \quad (5a)$$

where, if the variables are independent

$$\frac{k\sin\theta}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}=\frac{c\cos(\alpha+\theta)}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}=\frac{m\bar{X}-\bar{Y}}{\sqrt{(m^2\sigma_x^2+\sigma_y^2)}},$$

so that (5) is identical with Geary's formula.

Continuing to regard θ as the variable, the equation to the curve for the frequency distribution given by (5) is the moment of the area (4) about K, and this can be arrived at either geometrically or by differentiating (5) with regard to θ; it is

$$y = \frac{1}{\sqrt{(2\pi)}}\left\{\frac{k(a^2\sin\alpha\sin(\alpha+\theta)+b^2\cos\alpha\cos(\alpha+\theta))}{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))^{\frac{3}{2}}}\right\}$$
$$\times\exp\left[-\frac{1}{2}\left\{\frac{k\sin\theta}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}\right\}^2\right]. \quad (6)$$

Here the numerator of the factor within brackets may be put in the form

$$k\sqrt{(a^4\sin^2\alpha+b^4\cos^2\alpha)}\sin(\alpha+\gamma+\theta), \quad (7)$$

where γ is the absolute value of the angle which the axis conjugate to POP' makes with the major axis of the ellipse, i.e. where $\tan\gamma = (b^2/a^2)\cot\alpha$. It is thus seen that the curve is limited, the range of θ being from $-(\alpha+\gamma)$ to $\pi-(\alpha+\gamma)$; at these angles the value of y is zero.

It should be noted that the content between the planes SOS' and TKT' is equal to the content of the angle $P'KT'$ minus the content of the angle PKT, so that in the integration the value of the content of the angle PKT is twice neglected. It follows that if we integrate (6) between the limiting values of θ the total amount neglected is twice that part of the surface which lies beyond a plane GKG' which passes through and makes an angle of γ with the major axis, the value is

$$\frac{2}{\sqrt{(2\pi)}}\int_{k/\sigma_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\,dt,$$

where σ_α is the standard deviation of the normal curve given by a plane section of the normal surface which makes an angle of α with the major axis, i.e. where

$$\sigma_\alpha = \frac{ab}{\sqrt{(a^2\sin^2\alpha+b^2\cos^2\alpha)}}.$$

(3) CURVES GIVEN BY THIS EQUATION

The curves generated by this equation are of very great variety, the majority being bell-shaped, and we may now discuss the effects produced by varying the constants. It should be noted that, as POP' bisects the normal surface, the origin of the curve is neither the mean nor the mode but the median.

"k" may have any value from $3\sigma_\alpha$ or $4\sigma_\alpha$ up to infinity; as k tends to infinity the distribution of the standardized variable θ tends to normality with a standard deviation of

$$\frac{1}{k}\sqrt{(a^2\sin^2\alpha+b^2\cos^2\alpha)}.$$

As k decreases in value the departure from the form of the normal curve becomes more marked.

"a/b" may have any value from unity to infinity. When the value is unity the curve is very similar in form to the normal curve; as the value of a/b increases departure from the form of the normal curve increases.

On the value of α the symmetry of the curve depends; α may have any value between 0 and $\frac{1}{2}\pi$. At zero the curve is symmetrical but steeper than the normal curve. As α increases asymmetry develops (asymmetry being taken to mean the excentricity of the median), the maximum excentricity being reached at a value of $\tan^{-1}(b/a)$, thereafter the curve returns gradually to symmetry at a value of

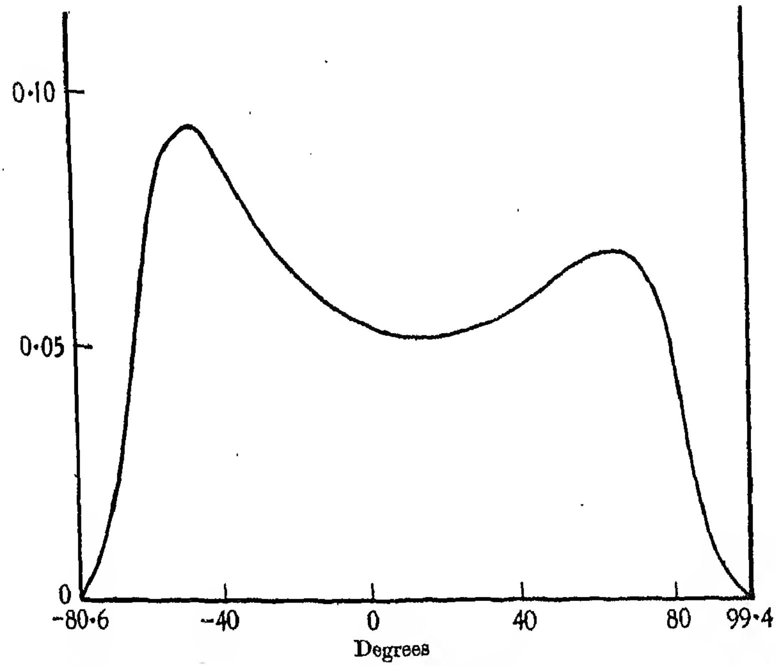


Fig. 2. Curve from equation (6). Constants: $a=4$, $b=1$, $k=3\sigma_\alpha$, $\alpha=80°$.

$\frac{1}{2}\pi$, where it is flatter than the normal curve. Skewness develops with asymmetry but more slowly, and the maximum skewness is not reached until α has a value of $\frac{1}{4}\pi$.

If k is relatively small, i.e. is near $3\sigma_\alpha$, and a/b is large, and if α is near $\frac{1}{2}\pi$, we get a curve with a maximum value on each side of the median, symmetrical when α is $\frac{1}{2}\pi$. This distribution does occasionally arise in practice; an example is given by Udny Yule (1932). Fig. 2 is an example of a slightly asymmetrical curve of this type.

(4) A general solution

K may very well occupy a position within the bulk of the surface; this must happen when both of the variables have negative values If we can obtain an equation of the curve in this case, it should have a general application to all values

of k. Geary stated this problem but did not proceed to its solution, Fieller gave a general algebraical solution.

We may consider K as lying on the circumference of the ellipse

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=h^2,$$

where $h=k/\sigma_\alpha$ and the ordinate on the circumference of the ellipse is

$$\frac{1}{2\pi ab}e^{-\frac{1}{2}h^2}.$$

In terms of the primary variables,

$$h=\frac{1}{\sqrt{(1-r^2)}}\left(\frac{\overline{X}^2}{\sigma_x^2}-\frac{2r\overline{X}\overline{Y}}{\sigma_x\sigma_y}+\frac{\overline{Y}^2}{\sigma_y^2}\right).$$

As before, the frequency curve of the angle θ is given by the positive moment of the normal curve (4) about the ordinate at K. This moment may be considered in two parts, the moment arising from the portion of the curve without the ellipse (A), and the moment arising from the portion of the curve within the ellipse on which K lies (B).

(A) For any normal curve the moment about the origin for that part of the curve beyond the ordinate at a given deviation $h\sigma$ is

$$y_0\int_{h\sigma}^{\infty}xe^{-\frac{1}{2}(x/\sigma)^2}dx,$$

which is equal to $$y_0\sigma^2e^{-\frac{1}{2}h^2}.$$

That is to say the moment is a function of the ordinate at the given deviation and of the standard deviation. Moreover the moment about this ordinate of the two equal tails of the curve is equal to the moment of these tails about the origin of the curve, so that in our case the required moment is

$$\frac{e^{-\frac{1}{2}h^2}}{\pi}\frac{ab}{a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta)}. \tag{8}$$

(B) The length of a chord of the ellipse which passes through K and makes an angle of $\alpha+\theta$ with the major axis is

$$\frac{2hab}{\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}\sqrt{\left\{1-\left(\frac{k\sin\theta}{h\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}\right)^2\right\}}, \tag{9}$$

and if we make $$\sin\phi=\frac{k\sin\theta}{h\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}} \tag{10}$$

we may put (9) as $2h\cos\phi\sigma_{(\alpha+\theta)}$. The total area of the normal curve in this plane using (4) is

$$\frac{1}{\sqrt{(2\pi)}\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}e^{-\frac{1}{2}(h\sin\phi)^2},$$

so that the ordinate at its origin is

$$\frac{1}{2\pi ab}e^{-\frac{1}{2}(h\sin\phi)^2},$$

and the area of the portion within the ellipse is

$$\frac{1}{\pi ab}e^{-\frac{1}{2}(h\sin\phi)^2}\int_0^{h\cos\phi\sigma_{(\alpha+\theta)}}\exp\left[-\frac{1}{2}\left(\frac{v}{\sigma_{(\alpha+\theta)}}\right)^2\right]dv. \tag{11}$$

This is multiplied by $h\cos\phi\sigma_{(\alpha+\theta)}$,

$$\frac{1}{\pi ab}e^{-\frac{1}{2}(h\sin\phi)^2}\sigma^2_{(\alpha+\theta)}h\cos\phi\int_0^{h\cos\phi}e^{-\frac{1}{2}u^2}du, \tag{12}$$

which may be simplified into

$$\frac{e^{-\frac{1}{2}h^2}}{\pi}\frac{ab}{a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta)}\left\{(h\cos\phi)^2+\frac{1}{1.3}(h\cos\phi)^4+\frac{1}{1.3.5}(h\cos\phi)^6+\frac{1}{1.3.5.7}(h\cos\phi)^8+\ldots\right\}; \tag{12a}$$

adding (8) and (12a), we have

$$y=\frac{e^{-\frac{1}{2}h^2}}{\pi}\frac{ab}{a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta)}\left\{1+(h\cos\phi)^2+\frac{1}{1.3}(h\cos\phi)^4+\frac{1}{1.3.5}(h\cos\phi)^6+\frac{1}{1.3.5.7}(h\cos\phi)^8+\ldots\right\} \tag{13}$$

as the equation of the curve. This expression converges quite rapidly for values of h up to 3; thereafter convergence becomes very slow indeed.

Reverting to (10), the value of $\sin\phi$ may be put in this form

$$\frac{ab\sin\theta}{\sqrt{(a^2\sin^2\alpha+b^2\cos^2\alpha)}\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}. \tag{10a}$$

From this the following identities may be established:

$$\cos\phi=\frac{a^2\sin\alpha\sin(\alpha+\theta)+b^2\cos\alpha\cos(\alpha+\theta)}{\sqrt{(a^2\sin^2\alpha+b^2\cos^2\alpha)}\sqrt{(a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta))}}, \tag{14}$$

$$\frac{d\phi}{d\theta}=\frac{ab}{a^2\sin^2(\alpha+\theta)+b^2\cos^2(\alpha+\theta)}, \tag{15}$$

$$\phi=\tan^{-1}\{(a/b)\tan(\alpha+\theta)\}-\tan^{-1}\{(a/b)\tan\alpha\}. \tag{16}$$

It should be noted that the limit of the integral in (5a) is $h\sin\phi$ and that the value of ϕ in (16) makes the practical work of calculating a series of these limits very simple. If the distribution of θ is still to be regarded as round the median at α, θ will have the limits as before, $-(\alpha+\gamma)$ and $\pi-(\alpha+\gamma)$, and at these limits ϕ will be $\mp\frac{1}{2}\pi$; generally, however, θ may be regarded as having a range of π beginning at any value, in which case ϕ will have the same range but with a different distribution.

A geometrical construction to show the relationship between θ and ϕ is given in Fig. 3, where OA and OB are equal to a and b respectively, and $O'AB$ is a right-angled isosceles triangle on AB as hypotenuse.

Here
$$AP = \frac{a \sin \alpha}{\sin OPA}$$

and
$$PB = \frac{b \sin\left(\frac{1}{2}\pi - \alpha\right)}{\sin OPB},$$

so that
$$AP/PB = (a/b) \tan \alpha;$$

by the same reasoning
$$AP/PB = \tan AO'P,$$

similarly
$$AO'Q = \tan^{-1}\{(a/b)\tan(\alpha+\theta)\};$$

therefore
$$PO'Q = \phi.$$

This relationship between θ and ϕ shows that the solution consists essentially in referring an asymmetrical system to an equivalent system where the standard deviations are equal.

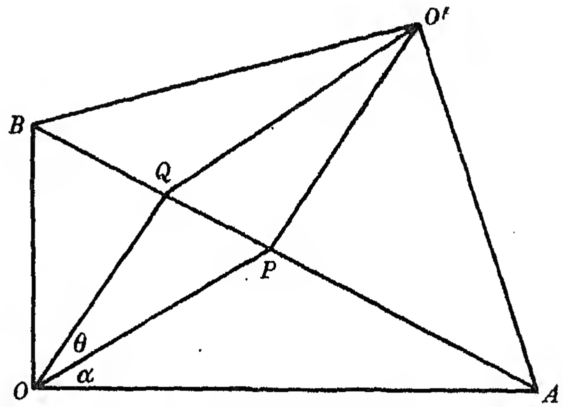


Fig. 3. Diagram to illustrate the relationship between θ and ϕ.

If we now turn back to (6), this can be put in the form

$$y = \frac{1}{\sqrt{(2\pi)}} \frac{hab\,(a^2 \sin\alpha \sin(\alpha+\theta) + b^2 \cos\alpha \cos(\alpha+\theta))}{(a^2 \sin^2\alpha + b^2\cos^2\alpha)^{\frac{1}{2}}\,(a^2\sin^2(\alpha+\theta) + b^2\cos^2(\alpha+\theta))^{\frac{3}{2}}}\, e^{-\frac{1}{2}(h \sin\phi)^2}, \quad (17)$$

so that we have the approximation of (6) to (13)

$$\frac{e^{-\frac{1}{2}h^2}}{\sqrt{(2\pi)}} \frac{d\phi}{d\theta} h\cos\phi\, e^{\frac{1}{2}(h\cos\phi)^2} \approx \frac{e^{-\frac{1}{2}h^2}}{\pi} \frac{d\phi}{d\theta} \left(1 + h\cos\phi\, e^{\frac{1}{2}(h\cos\phi)^2} \int_0^{h\cos\phi} e^{-\frac{1}{2}u^2} du\right). \quad (18)$$

It will be seen that the last expression may be written

$$\frac{e^{-\frac{1}{2}h^2}}{\sqrt{(2\pi)}} \frac{d\phi}{d\theta} h\cos\phi\, e^{\frac{1}{2}(h\cos\phi)^2} + \frac{e^{-\frac{1}{2}h^2}}{\pi} \frac{d\phi}{d\theta} \left(1 - h\cos\phi\, e^{\frac{1}{2}(h\cos\phi)^2} \int_{h\cos\phi}^{\infty} e^{-\frac{1}{2}u^2} du\right). \quad (19)$$

The difference between the values of y given by the two equations (if we do not take into account the value of $d\phi/d\theta$) is given in the table below for different values of h and for $\phi = \frac{1}{2}\pi$ and 0, showing that (6) must be a poor approximation to (13)

when h is less than 3, but that beyond that figure the difference rapidly becomes negligible.

h	Difference between (6) and (13)	
	$\phi = \frac{1}{2}\pi$	$\phi = 0$
1	0·193665	0·066475
2	0·043084	0·006776
3	0·003536	0·000305
4	0·000107	0·000005

Fieller's general solution of the problem is given in his formula (24)

$$\psi(v) = \frac{1}{\pi}\frac{\sigma_x\sigma_y\sqrt{(1-r^2)}}{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2}\exp\left[-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{\bar{x}^2}{\sigma_x^2} - 2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y} + \frac{\bar{y}^2}{\sigma_y^2}\right)\right]$$

$$+\exp\left[-\frac{1}{2}\frac{(\bar{y}-v\bar{x})^2}{\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2}\right]$$

$$\times\frac{\sigma_y(r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x(r\bar{x}\sigma_y - \bar{y}\sigma_x)}{\pi(\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2)^{\frac{3}{2}}}\int_0^{\frac{\sigma_y(r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x(r\bar{x}\sigma_y - \bar{y}\sigma_x)}{\sigma_x\sigma_y\{(1-r^2)(\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2)\}^{\frac{1}{2}}}} e^{-\frac{1}{2}u^2}\,du.$$

v is the value of the ratio under consideration which may be put as $\tan(\alpha+\delta+\theta)$, and if we reduce Fieller's formula with its symbols based on the co-ordinates of the primary variables to the system with its symbols based on the principal axes, we get the following identities:

$$\sigma_y^2 - 2rv\sigma_x\sigma_y + v^2\sigma_x^2 = (1+v^2)(a^2\sin^2(\alpha+\theta) + b^2\cos^2(\alpha+\theta)),$$

$$1 - r^2 = (a^2b^2)/(\sigma_x^2\sigma_y^2),$$

$$\frac{\bar{x}^2}{\sigma_x^2} - 2r\frac{\bar{x}}{\sigma_x}\frac{\bar{y}}{\sigma_y} + \frac{\bar{y}^2}{\sigma_y^2} = \frac{k^2(a^2\sin^2\alpha + b^2\cos^2\alpha)}{\sigma_x^2\sigma_y^2},$$

$$(\bar{y} - v\bar{x})^2 = (1+v^2)\,k^2\sin^2\theta,$$

$$\sigma_y(r\bar{y}\sigma_x - \bar{x}\sigma_y) + v\sigma_x(r\bar{x}\sigma_y - \bar{y}\sigma_x) = k\sqrt{(1+v^2)}$$
$$\times(a^2\sin\alpha\sin(\alpha+\theta) + b^2\cos\alpha\cos(\alpha+\theta)),$$

so that Fieller's formula as a whole reduces to

$$\psi(v) = \frac{1}{\pi}\frac{1}{1+v^2}\frac{d\phi}{d\theta}e^{-\frac{1}{2}h^2} + e^{-\frac{1}{2}(h\sin\phi)^2}\frac{d\phi}{d\theta}\frac{h\cos\phi}{\pi(1+v^2)}\int_0^{h\cos\phi} e^{-\frac{1}{2}u^2}\,du.$$

Here $d\phi/d\theta = 1+v^2$, so that the distribution $\psi(v)$ of v given by Fieller's equation is equivalent to the distribution of θ in the sum of (8) and (12).

The curves arising from (13) are of very great variety of form. They are all limited, indeed they are more properly described as cyclical, the ordinates at the limits for θ, since ϕ then equals $\mp\frac{1}{2}\pi$, being both equal to

$$\frac{e^{-\frac{1}{2}h^2}}{\pi}\frac{a^4\sin^2\alpha + b^4\cos^2\alpha}{ab(a^2\sin^2\alpha + b^2\cos^2\alpha)}. \tag{20}$$

That being so there is no necessity to regard the curve as beginning and ending at the limiting values for θ; in practice it would probably be taken to begin either where $\alpha+\theta$ is zero or where the ratio under consideration has a value of $-\infty$, i.e. where $\alpha+\delta+\theta$ is $-\frac{1}{2}\pi$. The unit deviation in both equations (6) and (13) is the radian; for practical work π/n would probably be used which would require a corresponding alteration in the equations. If we continue to regard the curve as

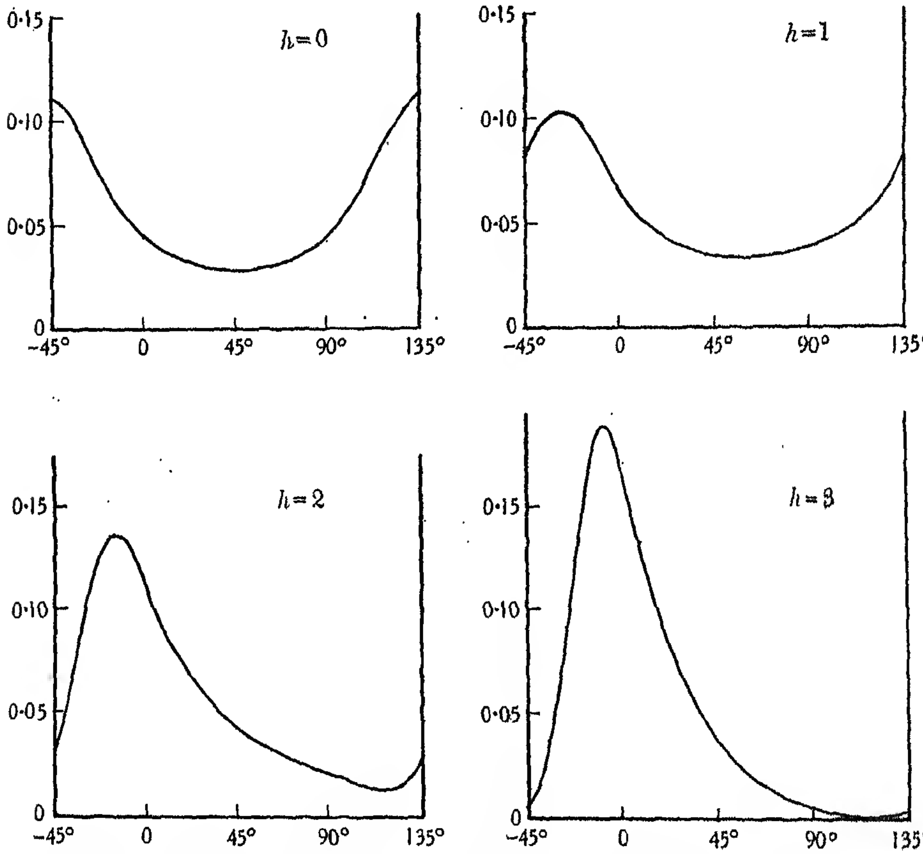


Fig. 4. Four curves from equations (13) to illustrate the change in form as h increases from 0 to 3. The curves commence where $\alpha+\theta=0$. Constants: $a=2$, $b=1$, $\alpha=45°$.

drawn with the median at α, we find that when h is zero it is U-shaped except when α is small, and is more or less asymmetrical depending on the value of α. With values of h about 3 it is possible to produce curves of a highly asymmetrical type; as h increases beyond 3 the curve reverts to the bell-shape and the limits recede from the bulk of the curve until, as we saw, when h is infinite the curve becomes the normal curve and is unlimited. Fig. 4 is an example of the development of the curve as h increases from zero to 3.

(5) Integration

The value of a frequency is the integration of y in (13) with respect to the variable θ and, making use of (15), this is

$$Q(\phi) = \frac{e^{-\frac{1}{2}h^2}}{\pi}\int\Big\{1 + (h\cos\phi)^2 + \frac{1}{1.3}(h\cos\phi)^4$$
$$+\frac{1}{1.3.5}(h\cos\phi)^6 + \frac{1}{1.3.5.7}(h\cos\phi)^8 + \ldots\Big\}\,d\phi. \quad (21)$$

An expression for this integral may be obtained by converting the series in powers of $\cos\theta$ into a series of cosines of multiples of ϕ, integration then gives a series in sines of multiples of ϕ; but the functions of h which are the factors in the series are very complicated and do not lend themselves to easy computation so that it is better to reconvert into a series in powers of $\cos\phi$, and (putting m for $\frac{1}{2}h^2$) this is

$$Q(\phi) = \frac{\phi}{\pi} + \frac{\sin\phi\cos\phi}{\pi}\Big\{(1-e^{-m}) + \tfrac{2}{3}(1-e^{-m}-e^{-m}m)\cos^2\phi$$
$$+\frac{2.4}{3.5}\left(1-e^{-m}-e^{-m}m-e^{-m}\frac{m^2}{2!}\right)\cos^4\phi$$
$$+\frac{2.4.6}{3.5.7}\left(1-e^{-m}-e^{-m}m-e^{-m}\frac{m^2}{2!}-e^{-m}\frac{m^3}{3!}\right)\cos^6\phi + \ldots\Big\}, \quad (22)$$

where the occurrence of the terms of Poisson's series is very interesting.

(6) Conclusion

The frequency distribution of the quotient of two normal variables may, then, give rise to most of the forms which are met with in statistical work. It is not, however, suggested that such statistical distributions always arise in this precise fashion; at the same time, from geometrical considerations, it seems likely that the product of two variables would produce a similar set of curves. It is not impossible that a large number of primary variables might group themselves into two secondary variables of approximately normal distribution and that the final distribution is some function involving either the quotient or the product of these variables. However that may be, the fitting of this curve to any given distribution appears to present many difficulties and is quite beyond the scope of this paper.

(7) Example

The following example illustrates the practical use of ϕ in the application of Geary's approximation. Some of the difficulties of childbirth are undoubtedly due to a disproportion between the size of the foetal head and the size of the bony opening through which it has to pass, the brim of the pelvis. This difficulty

becomes absolute, i.e. demands caesarian section, in about 1 % of all cases (in Guy's Hospital (1937) ten cases were dealt with by caesarian section on account of disproportion out of 990 pregnancies in a fair sample of the population); it would be well to know for the purposes of prognosis the percentage ratio between the size of the passenger and the size of the passage beyond which spontaneous delivery becomes impossible. The foetal head, as it passes, is roughly circular in section and the area of the maximum section may be calculated from the biparietal diameter; the figures for this diameter are taken from a series of 1010 measurements by Ince (1939). It has been shown that the area of the pelvic opening is given to a close enough approximation by the area of the ellipse on its antero-posterior (conjugate) and transverse diameters; the figures for these are taken from a series of 350 measurements made by radiology (Nicholson, 1938). It might be well to add that the radiological method used (Nicholson, 1936) has a probable error of accuracy as low as a millimetre.

These figures are

	Biparietal	Conjugate	Transverse
Mean (mm.)	91·5	116·4	132·3
Standard deviation (mm.)	4·0	10·5	7·6
Coefficient of variation	4·4	9·0	5·8

The distribution of these variables is normal, and the two latter are independent, so that we can estimate the following figures for the two areas:

	Foetal head (y)	Maternal pelvis (x)
Mean (sq. cm.)	65·8	121·0
Standard deviation (sq. cm.)	5·8	12·9
Coefficient of variation	8·8	10·7

The distribution of these variables is not, theoretically, normal but the error from assuming normality will be negligible; we shall also assume that they are independent, an assumption which is apparently not unreasonable. We may now calculate the following constants for the frequency curve for the ratio:

$$a = 12{\cdot}9; \quad b = 5{\cdot}8; \quad \alpha = \tan^{-1}(\bar{Y}/\bar{X}) = \tan^{-1}(65{\cdot}8/121{\cdot}0) = 28^\circ\, 32'{\cdot}25;$$
$$k = 137{\cdot}703; \quad \sigma_\alpha = 9{\cdot}357; \quad h = 14{\cdot}720.$$

From the tables of the normal curve we get the deviate for a frequency of 1 % as 2·3263, and applying (5a), (10), and (16), we have

$$h \sin\phi = 2{\cdot}3263,$$
$$\sin\phi = 0{\cdot}15804,$$
$$\phi = 9^\circ\, 6',$$
$$\tan^{-1}\{(a/b)\tan\alpha\} = 50^\circ\, 25',$$
$$\tan^{-1}\{(a/b)\tan(\alpha+\theta)\} = 59^\circ\, 31',$$
$$(a/b)\tan(\alpha+\theta) = 1{\cdot}69879,$$
$$\tan(\alpha+\theta) = 0{\cdot}764.$$

The required percentage ratio is then 76·4 %. Using the usual approximation $(\bar{Y}/\bar{X})\{1+(\sigma_x/\bar{X})^2\}$, the mean of the ratio would be 55·5 %, and its standard deviation $(\bar{Y}/\bar{X})\{(\sigma_x/\bar{X})^2+(\sigma_y/\bar{Y})^2\}$, 7·5 %; if we had assumed that the distribution of the ratio was normal, we should have got a result of 72·9 %; so that, even when h is quite high, the distribution of the tails of the curve may be far from normal.

The value of the figure 76·4 from the point of view of prognosis is that we can now predict that unless an event has occurred, the chances against which are 99 to 1, a pelvis with an area of 110 sq. cm. can pass 99·9 % of foetal heads, that a pelvis of 100 sq. cm. can pass 97 %, that a pelvis of 90 sq. cm. can pass 70 %, but that a pelvis of 80 sq. cm. can pass no more than 21 % of foetal heads.

REFERENCES

Fieller, E. C. (1932). *Biometrika*, **24**, 428.
Geary, R. C. (1930). *J. R. Statist. Soc.* **93**, 442.
Guy's Hospital (1937). *Clinical Report of the Maternity Department.*
Ince, J. G. Hastings (1939). *J. Obstet. Gynaec.* **46**, 1003.
Nicholson, C. (1936). *Lancet*, no. 231, 615.
—— (1938). *J. Obstet. Gynaec.* **45**, 950.
Udny Yule, G. (1932). *Theory of Statistics*, 10th ed. London: Griffin, 78.

THE STATISTICAL SIGNIFICANCE OF CANONICAL CORRELATIONS

By M. S. BARTLETT

1. In an important paper published in this *Journal*, Hotelling (1936) has shown that the generalized variance matrix*

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

of a vector variate $\mathbf{x}$ which has been partitioned into two parts $\mathbf{x}_1$ and $\mathbf{x}_2$ with, say, q and p components, can, by appropriate linear transformations $\mathbf{L}_1\mathbf{x}_1$ and $\mathbf{L}_2\mathbf{x}_2$ of $\mathbf{x}_1$ and $\mathbf{x}_2$, be thrown into the canonical form

$$\begin{pmatrix} \mathbf{L}_1\mathbf{V}_{11}\mathbf{L}_1' & \mathbf{L}_1\mathbf{V}_{12}\mathbf{L}_2' \\ \mathbf{L}_2\mathbf{V}_{21}\mathbf{L}_1' & \mathbf{L}_2\mathbf{V}_{22}\mathbf{L}_2' \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{R} \\ \mathbf{R} & 1 \end{pmatrix}.$$

$\mathbf{R}$ is a rectangular matrix which is zero except for a leading diagonal of squares λ_i^2 of canonical correlations.

Similar operations on the estimated matrix variance V give rise to estimated canonical correlations l_i, which measure the correlations between estimates of the linear functions $\mathbf{L}_1\mathbf{x}_1$ and $\mathbf{L}_2\mathbf{x}_2$. While Hotelling has given asymptotic standard errors for the coefficients l_i, it is known that the significance of these correlations, as in the simple case $p = 1$, is more generally to be interpreted as the significance of the regression relations of $\mathbf{x}_2$ with $\mathbf{x}_1$, the validity of any exact tests of significance depending on the supposition that the dependent variate $\mathbf{x}_2$, apart from its linear dependence on $\mathbf{x}_1$, is normal.

Special cases of the simultaneous distribution of the correlations l_i, when $\mathbf{x}_2$ and $\mathbf{x}_1$ are unrelated, have been considered by Hotelling (1936) and Girschick (1939), but an important theoretical advance is represented by the derivation of the distribution (under the same conditions) for any values of p and q (Fisher, 1939; Hsu, 1939). It will be shown that this distribution makes available further possible tests; and since the problem of the most appropriate tests of significance

* A matrix is usually denoted by a capital letter, and if it has both a population and sample value, the population value is given in heavier type (cf. Bartlett, 1939). The transpose of any matrix A is denoted by A'. A matrix with only one column is a vector, and is often denoted by a small letter. To avoid confusion, a vector variate $\mathbf{x}$ is written in heavier type throughout, to distinguish it from a single variate x. If $\mathbf{x}$ is measured from its population mean, the variance matrix $\mathbf{V}$ is the average value of $\mathbf{x}\mathbf{x}'$. In practice we lose one or more degrees of freedom by measuring $\mathbf{x}$ from sample or regression means, but without loss of generality we shall suppose that our sample consists of measurements of $\mathbf{x}$ with ν degrees of freedom.

has not always been considered very adequately by other writers, it is also the purpose of this paper to explain the logical relation of these further tests to tests of significance previously available.*

2. For the case $p \leqslant q$, the distribution of l_i, when these roots are arranged in order of magnitude, is given by

$$F(l_1^2, l_2^2, \ldots, l_p^2)\, dl_1^2\, dl_2^2 \ldots dl_p^2,$$

where
$$F = C \prod_{i=1}^{p} \left\{ (l_i^2)^{\frac{1}{2}(q-p-1)} (1-l_i^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{j=i+1}^{p} (l_i^2 - l_j^2) \right\}$$

and
$$C = \pi^{\frac{1}{2}p} \prod_{i=1}^{p} \frac{\Gamma\frac{1}{2}(\nu-i+1)}{\Gamma\frac{1}{2}(q-i+1)\,\Gamma\frac{1}{2}(\nu-q-i+1)\,\Gamma\frac{1}{2}(p-i+1)}. \qquad (1)$$

For $p > q$, we need only reverse the roles of $\mathbf{x}_1$ and $\mathbf{x}_2$.

A criterion which is useful in detecting the simultaneous departure of several roots λ_i from zero is the product

$$\prod_{i=1}^{p} (1-l_i^2) = \Lambda, \quad \text{say.}\dagger$$

When $p = 1$, the distribution of Λ is equivalent to that of l_1^2, and the distribution in (1) can be transformed if required into Fisher's z-distribution. When $p = 2$, it was found by Wilks that a similar distribution exists for $\sqrt{\Lambda}$. For $p > 2$, no exact test is at present available, but the formula

$$\chi^2 = -\{\nu - \tfrac{1}{2}(p+q+1)\} \log \Lambda,$$

with pq degrees of freedom, gives a good approximate test (Bartlett, 1938).

If the roots $\lambda_2^2, \ldots, \lambda_p^2$ are zero, we are, however, including in Λ irrelevant degrees of freedom which might possibly obscure the significance of λ_1^2. For any test on λ_1^2 by itself, we have little choice but to consider l_1^2, though we do not really know whether l_1^2 is the root corresponding to λ_1^2 or not. The probability distribution‡ $p(l_1^2)$ is theoretically obtainable from (1), and hence also 0·05 or 0·01 levels

* The distribution of l_i^2 obtained by Fisher and Hsu has also been obtained by Roy (1939), though this writer was concerned with the different problem of comparing the dispersion in two multivariate normal samples. For a single variate, testing the significance of a sum of squares separated off from the total sum of squares by a multiple correlation or regression formula is equivalent to testing the ratio of two variances, a criterion also employed to test the equality of two population variances. Roy has proposed generalizing the latter problem along lines which give rise to the same distribution problem solved by Fisher and Hsu, but while he has independently obtained the same general distribution, the need for some care in the choice of tests in multivariate analysis is even more evident in the problem with which Roy was concerned. It is obvious, for example, that the p roots which Roy considered cannot represent *all* the possible differences among the $\frac{1}{2}p(p+1)$ variance and covariance parameters between two p-variate normal samples, and some explanation of their interpretation seems required.

† This criterion has been proposed by Wilks (1932), Bartlett (1934), and Hotelling (1936), the last-named denoting it by z.

‡ The probability of a random variable having a particular value x is denoted by $p(x)$. If the variable has a continuous range of values, $p(x)$ denotes the probability of the variable falling in the interval $x, x+dx$. The corresponding notations $x \mid y$ and $p(x \mid y)$ are used when the variable is only being considered for a fixed value y of another variable. The probability symbol p is not of course to be confused with the number p of components in the vector variate $\mathbf{x}_2$.

of significance of l_1^2 for specified values of ν, p and q. The tabulation of these levels would be useful, but would also be a task of some magnitude, and it is therefore worth noting that owing to the problem of identification, the largest root l_1^2 is not a sufficient statistic for λ_1^2, and $p(l_1^2)$ has no unique relevance. If we consider, instead, the distribution of l_1^2 for given values of $l_2^2 \dots l_p^2$, we have corresponding to the probability relation

$$p(l_1^2, \dots, l_p^2) = p(l_1^2 \mid l_2^2, \dots, l_p^2)\, p\,(l_2^2, \dots, l_p^2),$$

the probability density relation

$$F(l_1^2, \dots, l_p^2) = f_1(l_1^2 \mid l_2^2, \dots, l_p^2) f_2(l_2^2, \dots, l_p^2),$$

where the function f_1, apart from the constant term f_2, is determined at once from the function F.

In the logical situation we are postulating where λ_1^2, but not the other roots, is different from zero, it is not evident which distribution, $p(l_1^2)$ or $p(l_1^2 \mid l_2^2, \dots, l_p^2)$, provides the more powerful test, owing to the absence of sufficiency properties, and it is of some interest to consider in detail another problem which is trivial in itself, but serves to illustrate the principles involved.

3. Suppose we have a pair of variates x_1 and x_2 both independently following a rectangular distribution

$$p\,(x) = dx, \quad (0 \leqslant x \leqslant 1).$$

One variate (unspecified) is then shifted a distance α, so that it follows the distribution

$$p(x) = dx, \quad (\alpha \leqslant x \leqslant 1+\alpha).$$

If x_1 and x_2 denote the variates in order of magnitude, we shall detect the shift α from the larger value, x_1, if α is large enough. To compare the value of $p(x_1)$ and $p(x_1 \mid x_2)$, we note first of all that when $\alpha = 0$,

$$p(x_1) = 2x_1\,dx_1, \quad (0 \leqslant x_1 \leqslant 1)$$

$$p(x_1 \mid x_2) = \frac{dx_1}{1-x_2}. \quad (x_2 \leqslant x_1 \leqslant 1)$$

For the significance level ϵ, $p(x_1)$ gives a critical value $x_1 = \sqrt{(1-\epsilon)}$, while $p(x_1 \mid x_2)$ gives $x_1 = 1-\epsilon(1-x_2)$. If α is different from zero, a peculiar feature (analogous to the canonical correlation problem) is that the larger observation x_1 may or may not be associated with α. For $p(x_1)$ we find

$$(2x_1 - \alpha)\,dx_1, \quad (\alpha \leqslant x_1 \leqslant 1)$$

$$dx_1. \quad (1 < x_1 \leqslant 1+\alpha)$$

For $p(x_1 \mid x_2)$, we have

$$\frac{2dx_1}{\alpha + 2(1-x_2)}, \quad (x_2 \leqslant x_1 \leqslant 1)$$

$$\frac{dx_1}{\alpha + 2(1-x_2)}. \quad (1 < x_1 \leqslant 1+\alpha)$$

This is provided $x_2 \geqslant \alpha$; for $x_2 < \alpha$, we have

$$p(x_1 \mid x_2) = dx_1. \quad (\alpha \leqslant x_1 \leqslant 1+\alpha)$$

Using the terminology of Neyman and Pearson, we shall denote the power of the test derived from x_1 by P; and for that from $x_1 \mid x_2$, by P'. Then

$$1-P = \int_{\alpha}^{\sqrt{(1-\epsilon)}} (2x_1-\alpha)\,dx_1$$
$$= 1-\epsilon-\alpha\sqrt{(1-\epsilon)}.$$

For $1-P'$, we have first of all, for given x_2, an integral

$$\int_{x_2}^{1-\epsilon(1-x_2)} p(x_1 \mid x_2),$$

which gives

$$1-\epsilon(1-x_2)-\alpha, \quad (x_2<\alpha)$$
$$\frac{2(1-\epsilon)(1-x_2)}{\alpha+2(1-x_2)}. \quad (x_2 \geqslant \alpha)$$

Since $p(x_2 \mid \alpha)$ is given by

$$\{\alpha+2(1-x_2)\}\,dx_2, \quad (\alpha \leqslant x_2 \leqslant 1)$$
$$dx_2, \quad (0 \leqslant x_2 < \alpha)$$

we finally obtain, after averaging over x_2,

$$1-P' = (1-\epsilon)(1-\alpha)-\tfrac{1}{2}\alpha^2\epsilon.$$

Before comparing P with P', we may remember that we do not expect either x_1 or $x_1 \mid x_2$ to provide the most powerful test obtainable. Theoretically we can see what this test would be by considering the ratio

$$p(x_1, x_2 \mid \alpha)/p(x_1, x_2 \mid 0) = X_\alpha,$$

say, though since the value of X_α is indeterminate unless the true value of α is specified, it should be realized that X_α does not provide us with any actual test, only with a theoretical upper limit for P or P'.

The criterion X_α has the distribution

$$\begin{array}{llll} X_\alpha = & \infty & 1 & 0 \\ p(X_\alpha \mid \alpha) = & \alpha & (1-\alpha)^2 & \alpha(1-\alpha) \\ p(X_\alpha \mid 0) = & 0 & (1-\alpha)^2 & 2\alpha-\alpha^2 \end{array}$$

For $(1-\alpha)^2 \geqslant \epsilon$, we shall allow the value $X_\alpha = 1$ to be significant in $\epsilon/(1-\alpha)^2$ of the times that the value 1 occurs; if $(1-\alpha)^2 \leqslant \epsilon$, we allow $X_\alpha = 0$ to be significant in the fraction

$$\frac{\epsilon-(1-\alpha)^2}{2\alpha-\alpha^2}$$

of the times that $X_\alpha = 0$ occurs. The power P'' of a test that could be based on X_α is then

$$\alpha+\epsilon, \quad [(1-\alpha)^2 \geqslant \epsilon],$$
$$\alpha+(1-\alpha)^2+\left\{\frac{\epsilon-(1-\alpha)^2}{2\alpha-\alpha^2}\right\}\alpha(1-\alpha), \quad [(1-\alpha)^2 \leqslant \epsilon].$$

Comparative values of P, P' and P'' are given in Table 1 for $\epsilon = 0{\cdot}05$ and $0{\cdot}10$.

Table 1

ϵ	α	0	0·1	0·2	0·4	0·6	0·8	0·9
0·05	P	0·0500	0·1475	0·2449	0·4399	0·6348	0·8298	0·9272
	P'	0·0500	0·1453	0·2410	0·4340	0·6290	0·8260	0·9253
	P''	0·0500	0·1500	0·2500	0·4500	0·6500	0·8464	0·9373
0·10	P	0·1000	0·1949	0·2897	0·4795	0·6692	0·8590	0·9538
	P'	0·1000	0·1905	0·2820	0·4680	0·6580	0·8520	0·9505
	P''	0·1000	0·2000	0·3000	0·5000	0·7000	0·8785	0·9603

It will be seen that $p(x_1)$ provides a test in this problem rather more powerful than $p(x_1 \mid x_2)$, but that the latter is quite effective. We cannot of course transfer this result to our main problem, but it is clear that $p(l_1^2 \mid l_2^2, \ldots, l_p^2)$ may justifiably be considered, at least until the distribution $p(l_1^2)$ has been tabulated.

4. Returning then to this distribution, we may examine one or two special cases before formally noting the significance level of l_1^2 in general. It has been shown by Fisher and Hsu that for ν large, the distribution of $l_1^2, l_2^2, \ldots, l_p^2$ tends to

$$G(m_1^2, m_2^2, \ldots, m_p^2)\, dm_1^2\, dm_2^2 \ldots dm_p^2,$$

where $m_i^2 = \frac{1}{2}\nu l_i^2$, $\quad G = C' \prod_{i=1}^{p} \left\{ (m_i^2)^{\frac{1}{2}(q-p-1)}\, e^{-m_i^2} \prod_{j=i+1}^{p} (m_i^2 - m_j^2) \right\},$

and $$1/C' = \prod_{i=1}^{p} \{\Gamma\tfrac{1}{2}(q-i+1)\, \Gamma\tfrac{1}{2}(p-i+1)\}. \tag{2}$$

For the particular case $p = 2$, $q = 3$, the distribution of $m_1^2 \mid m_2^2$ is

$$(m_1^2 - m_2^2)\, e^{-(m_1^2 - m_2^2)}\, dm_1^2,$$

which is a function simply of $m_1^2 - m_2^2$. If alternatively we consider the distribution $p(m_1^2)$, we obtain

$$2e^{-m_1^2}\{e^{-m_1^2} - (1 - m_1^2)\}\, dm_1^2,$$

the 0·05 significance level for which is 5·37. From $p(m_1^2 \mid m_2^2)$ this value of 5·37 corresponds to a level 0·030 if $m_2^2 = 0$, to 0·045 if m_2^2 is equal to its expected value 0·50, and to 0·05 when m_2^2 reaches the value 0·63. These results merely illustrate how the significance level of m_1^2 depends on which distribution is being used.

For the case $p = 3$, $q = 4$, the significance level for m_1^2 can be written

$$e^{-u}\left\{(u+1) + \frac{u^2}{2+v}\right\},$$

where $u = m_1^2 - m_2^2$, $v = m_2^2 - m_3^2$. The level of significance thus depends mainly on u, as we should expect, but the effect of v is not negligible. The factor multiplying the exponential varies, for example, when $u = 4$, from $10\frac{1}{3}$ for $v = 1$ to 13 for $v = 0$.

The general expression for the significance level for m_1^2 is

$$\frac{\int_{m_1^2}^{\infty} (m_1^2)^{\frac{1}{2}(q-p-1)} e^{-m_1^2} \prod_{i=2}^{p} (m_1^2 - m_i^2)\, dm_1^2}{\int_{m_2^2}^{\infty} (m_1^2)^{\frac{1}{2}(q-p-1)} e^{-m_1^2} \prod_{i=2}^{p} (m_1^2 - m_i^2)\, dm_1^2},$$

or, if we write $$\Gamma_x(\alpha) = \int_x^{\infty} x^{\alpha-1} e^{-x}\, dx,$$

by

$$\frac{\Gamma_{m_1^2}(\tfrac{1}{2}[p+q+1]) - \left\{\sum_{i=2}^{p} m_i^2\right\} \Gamma_{m_1^2}(\tfrac{1}{2}[p+q-1]) + \left\{\sum_{i=2}^{p} \sum_{j=i+1}^{p} m_i^2 m_j^2\right\} \Gamma_{m_1^2}(\tfrac{1}{2}[p+q-3])}{\Gamma_{m_2^2}(\tfrac{1}{2}[p+q+1]) - \left\{\sum_{i=2}^{p} m_i^2\right\} \Gamma_{m_2^2}(\tfrac{1}{2}[p+q-1]) + \left\{\sum_{i=2}^{p} \sum_{j=i+2}^{p} m_i^2 m_j^2\right\} \Gamma_{m_2^2}(\tfrac{1}{2}[p+q-3])}. \tag{3}$$

For the more general case of finite ν, we have similarly for l_1^2,

$$\frac{\int_{l_1^2}^{1} (l_1^2)^{\frac{1}{2}(q-p-1)} (1-l_1^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{i=2}^{p} (l_1^2 - l_i^2)\, dl_1^2}{\int_{l_2^2}^{1} (l_1^2)^{\frac{1}{2}(q-p-1)} (1-l_1^2)^{\frac{1}{2}(\nu-q-p-1)} \prod_{i=2}^{p} (l_1^2 - l_i^2)\, dl_1^2},$$

or, if $$B_x(\alpha, \beta) = \int_x^{1} x^{\alpha-1} (1-x)^{\beta-1}\, dx,$$

by

$$\frac{B_{l_1^2}(\tfrac{1}{2}[p+q+1], \tfrac{1}{2}[\nu-p-q+1]) - \left\{\sum_{i=2}^{p} l_i^2\right\} B_{l_1^2}(\tfrac{1}{2}[p+q-1], \tfrac{1}{2}[\nu-p-q+1]) + \dots}{B_{l_2^2}(\tfrac{1}{2}[p+q+1], \tfrac{1}{2}[\nu-p-q+1]) - \left\{\sum_{i=2}^{p} l_i^2\right\} B_{l_2^2}(\tfrac{1}{2}[p+q-1], \tfrac{1}{2}[\nu-p-q+1]) + \dots}. \tag{4}$$

The dependence of results (3) and (4) not only on ν, p and q, but also on the particular values of $l_2^2, \dots, l_p^2$, makes it impracticable to tabulate the 0·05 or other levels of significance; but it is not difficult in any instance to find the exact level from (3) or (4), using the published tables of $\Gamma_x(\alpha)$ or $B_x(\alpha, \beta)$.*

It must be recognized that if the second root λ_2^2 is also different from zero, the distribution of l_1^2 for given l_2^2 is quite irrelevant, but except possibly when p is rather large, it is probable that two or more non-zero roots would be detected by the Λ criterion, and the testing of λ_1^2 alone by means of l_1^2 (a test which is still not completely efficient) would not arise.

5. Directly we have established the existence of at least one root λ_1^2, we may always proceed to eliminate this correlation λ_1 and the corresponding pair of canonical variates; and analyse the remainder. The theory of eliminating from $\mathbf{x}_2$ a set of specified variates represented, say, by the vector variate $\mathbf{x}_0$ has been

* *Tables of the Incomplete Γ-function*, ed. K. Pearson (1922, His Majesty's Stationery Office, London); *Tables of the Incomplete Beta-function*, ed. K. Pearson (1934, Biometrika Office, University College, London).

indicated by Bartlett (1939).* As a particular case, $\mathbf{x}_0$ may be a hypothetical set of r canonical variates of $\mathbf{x}_2$, and the criterion $\Lambda(\nu-r, p-r, q)$ for the remaining variate $\mathbf{x}_{2\cdot 0}$, in place of the original criterion $\Lambda(\nu, p, q)$ for $\mathbf{x}_2$, would test the goodness of fit of the hypothetical vector canonical variate $\mathbf{x}_0$. In the case $q = 1$, we have the goodness of fit of a hypothetical discriminant function, the problem of which was first raised by Fisher (1938).

It has, however, also been pointed out (Bartlett, 1938, p. 39) that if the canonical vector variate $\mathbf{x}_0$ has been estimated from the data, the symmetrical relation between $\mathbf{x}_2$ and $\mathbf{x}_1$ will imply that each has only $p-r$ and $q-r$ independent components remaining, the χ^2 approximation for the criterion

$$\Lambda' = \prod_{i=r+1}^{p} (1-l_i^2)$$

being $$-\{(\nu-r)-\tfrac{1}{2}[(p-r)+(q-r)+1]\}\log \Lambda' = -\{\nu-\tfrac{1}{2}(p+q+1)\}\log \Lambda',$$

with $(p-r)(q-r)$ degrees of freedom. It was stressed that this reduction of the degrees of freedom essentially depends on the existence of non-zero roots $\lambda_1^2, \ldots, \lambda_r^2$, so that the vector variate $\mathbf{x}_0$ is well-determined, and any effect of selection of $l_1^2, \ldots, l_r^2$ from $l_1^2, \ldots, l_p^2$ can be neglected. Under the same conditions, we may approximately use the tests known for $p = 1$ or 2, for the criterion $\Lambda'(\nu-r, p-r, q-r)$, when $p-r = 1$ or 2.

6. To demonstrate the reduction in degrees of freedom in the case $r = 1$, consider the case when ν is large, and

$$-\nu \log \Lambda \to \sum_{i=1}^{p} \nu l_i^2 \to \chi^2.$$

If $\nu l_i^2 = \theta_i$, the determinantal equation for θ_i is of the form

$$|A-\theta \mathbf{V}| = 0,$$

where $\mathbf{V}$ denotes the variance matrix of $\mathbf{x}_2$, and A is a matrix of the sums of squares and products among the p variates of $\mathbf{x}_2$ for that portion of the sample separated off in terms of the independent vector variate $\mathbf{x}_1$. Without loss of generality, we shall suppose that $\mathbf{V} = \mathbf{I}$.

Regarding the ν observations for any variate as a vector with ν orthogonal components, let us now add to the chance variation of the first variate of $\mathbf{x}_2$ a part dependent on each of the q (orthogonal) variates of $\mathbf{x}_1$. For each variate of $\mathbf{x}_1$, the length of the vector representing the first variate of $\mathbf{x}_2$ will then receive an addition X_k, say, $(k=1\ldots q)$, which will be of order $\sqrt{\nu}$. Partitioning off the first variate of $\mathbf{x}_2$, we obtain, as our new equation for θ,

$$\left|\begin{array}{c:c} a_{11}+2\Sigma x_1 X+\Sigma X^2-\theta & a_{1j}+\Sigma x_j X \\ \hdashline a_{i1}+\Sigma x_i X & a_{ij}-\theta \end{array}\right| = 0.$$

* See equation (2.8) of the paper cited, and the immediately preceding equation.

The summation sign is for the q degrees of freedom of $\mathbf{x}_1$, and $a_{ij} = \Sigma x_i x_j$, where $x_1, \ldots, x_p$ are the p variates of $\mathbf{x}_2$. Solving the equation for the largest root, we have

$$\theta_1 = \Sigma X^2 + 2\Sigma x_1 X + a_{11} + \sum_{i=2}^{p} \frac{(\Sigma x_i X)^2}{\Sigma X^2} + o\left(\frac{1}{\sqrt{(\Sigma X^2)}}\right).$$

If we neglect the last term, the sum of the remaining roots becomes

$$\sum_{i=1}^{p} \theta_i - \theta_1 = \sum_{i=2}^{p} \left\{a_{ii} - \frac{(\Sigma x_i X)^2}{\Sigma X^2}\right\},$$

which is a χ^2 with $(p-1)(q-1)$ degrees of freedom.

7. To illustrate the use of this test we may consider the data from Kelley quoted by Hotelling (1936), these consisting of correlations among tests in (1) reading speed, (2) reading power, (3) arithmetic speed and (4) arithmetic power, the sample being one of 140 seventh-grade school children. Hotelling, investigating the relation of arithmetical with reading abilities, found canonical correlations

$$l_1 = 0{\cdot}3945, \quad l_2 = 0{\cdot}0688.$$

Since $\nu = 139$, $p = 2$, $q = 2$, the first correlation gives a contribution to χ^2 of

$$-\{139 - \tfrac{1}{2}(2+2+1)\}\log(1 - 0{\cdot}3945^2) = 23{\cdot}09.$$

Similarly the contribution from l_2 is 0·64. The χ^2 analysis is consequently summarized as in Table 2.

Table 2

	D.F.	χ^2
l_1	3	23·09
l_2	1	0·64
Total	4	23·73

It is evident at once, as Hotelling concluded from other tests, that there is a significant relation between arithmetical and reading abilities, which arises entirely from the first canonical correlation.

REFERENCES

BARTLETT, M. S. (1934). "The vector representation of a sample." *Proc. Camb. Phil. Soc.* **30**, 327–40.

—— (1938). "Further aspects of the theory of multiple regression." *Proc. Camb. Phil. Soc.* **34**, 33–40.

—— (1939). "A note on tests of significance in multivariate analysis." *Proc. Camb. Phil. Soc.* **35**, 180–5.

FISHER, R. A. (1938). "The statistical utilization of multiple measurements." *Ann. Eugen.* **8**, 376–86.

—— (1939). "The sampling distribution of some statistics obtained from non-linear equations." *Ann. Eugen.* **9**, 238–49.

GIRSCHICK, M. A. (1939). "On the sampling theory of the roots of determinantal equations." *Ann. Math. Statist.* **10**, 203–24.

HOTELLING, H. (1936). "Relations between two sets of variates." *Biometrika*, **28**, 321–77.

HSU, P. L. (1939). "On the distribution of roots of certain determinantal equations." *Ann. Eugen.* **9**, 250–8.

ROY, S. N. (1939). "p-statistics or some generalizations in analysis of variance appropriate to multivariate problems." *Sankhyā*, **4**, 381–96.

WILKS, S. S. (1932). "Certain generalizations in the analysis of variance." *Biometrika*, **24**, 471–94.

ON THE LIMITING DISTRIBUTION OF THE CANONICAL CORRELATIONS

By P. L. HSU

1. The purpose of this paper is to deduce the limiting distribution of Hotelling's canonical correlations* under the most general assumption on the population canonical correlations. The result is stated in the theorem at the end of this paper.

The method employed here is essentially the same as that by which we derived (Hsu, 1940) the limiting distribution of Fisher's discriminating components. In what follows, steps in the derivation are given while strictly rigorous reasoning is left out. The latter may be found in the author's 1940 paper.

The parent distribution is represented by the density

$$\text{const.}\exp\left\{-\tfrac{1}{2}\left(\sum_{i,j=1}^{p}\alpha_{ij}a_{ij}+2\sum_{i=1}^{p}\sum_{g=1}^{q}\beta_{ig}b_{ig}+\sum_{g,h=1}^{q}\gamma_{gh}c_{gh}\right)\right\},\quad (p\leqslant q),\qquad \ldots\ldots(1)$$

where
$$a_{ij}=\sum_{t=1}^{n}x_{it}x_{jt},\quad b_{ig}=\sum_{t=1}^{n}x_{it}y_{gt},\quad c_{gh}=\sum_{t=1}^{n}y_{gt}y_{ht}.\qquad \ldots\ldots(2)$$

By virtue of Hotelling's reduction, the matrix of variances and covariances is taken to be

$$\begin{Vmatrix} \alpha_{11} & \cdots & \alpha_{1p} & \beta_{11} & \cdots & \beta_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{p1} & \cdots & \alpha_{pp} & \beta_{p1} & \cdots & \beta_{pq} \\ \beta_{11} & \cdots & \beta_{p1} & \gamma_{11} & \cdots & \gamma_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \beta_{1q} & \cdots & \beta_{pq} & \gamma_{q1} & \cdots & \gamma_{qq} \end{Vmatrix}^{-1} = \begin{Vmatrix} 1 & \ldots & 0 & \rho_1' & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \ldots & 1 & 0 & \ldots & \rho_p' & \ldots & 0 \\ \rho_1' & \ldots & 0 & 1 & \ldots & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \ldots & \rho_p' & 0 & \ldots & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & \ldots & 1 \end{Vmatrix}, \qquad \ldots\ldots(3)$$

where $\rho_1', \ldots, \rho_p'$ are the population canonical correlations. The sample canonical correlations, $r_1, \ldots, r_p$, are the positive roots of the equation†

$$\begin{vmatrix} ra_{11} & \cdots & ra_{1p} & b_{11} & \cdots & b_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ ra_{p1} & \cdots & ra_{pp} & b_{p1} & \cdots & b_{pq} \\ b_{11} & \cdots & b_{p1} & rc_{11} & \cdots & rc_{1q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{1q} & \cdots & b_{pq} & rc_{q1} & \cdots & rc_{qq} \end{vmatrix} = 0. \qquad \ldots\ldots(4)$$

* Hotelling (1936). For further work on the distribution of the canonical correlations, see Madow (1938), Girschick (1939) and Hsu (1939).

† We use r in (4) instead of $-r$ as in Hotelling's original definition because it is known that the non-vanishing roots of (4) form pairs each of which have the same absolute value but opposite signs.

We set

$$\left.\begin{aligned} \rho'_1 = \dots = \rho'_{\mu_1} &= \rho_1, \\ \rho'_{\mu_1+1} = \dots = \rho'_{\mu_1+\mu_2} &= \rho_2, \\ \dots\dots\dots\dots\dots\dots & \\ \rho'_{\mu_1+\dots+\mu_{\nu-1}+1} = \dots = \rho'_s &= \rho_\nu, \end{aligned}\right\} \qquad \dots\dots(5)$$

$$\rho'_{s+1} = \dots = \rho'_p = 0, \qquad \dots\dots(6)$$

$$\rho_1 > \rho_2 > \dots > \rho_\nu > 0, \qquad \dots\dots(7)$$

$$r_1 \geqslant r_2 \geqslant \dots \geqslant r_p, \qquad \dots\dots(8)$$

and proceed to find the limiting distribution of $r_1, \dots, r_p$ as $n \to \infty$.

2. Lemma 1. *We have the identity*

$$\begin{vmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{vmatrix} = |\mathbf{S}| . |\mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R}|. \qquad \dots\dots(9)$$

This results from the identity

$$\begin{Vmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{R} & \mathbf{S} \end{Vmatrix} . \begin{Vmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{S}^{-1}\mathbf{R} & \mathbf{I} \end{Vmatrix} = \begin{Vmatrix} \mathbf{P} - \mathbf{Q}\mathbf{S}^{-1}\mathbf{R} & \mathbf{Q} \\ \mathbf{O} & \mathbf{S} \end{Vmatrix}$$

on taking determinants on both sides.

Lemma 2. *Let*

$$\left.\begin{aligned} a_{ii} &= n + \sqrt{n}\, u_{ii}, & a_{ij} &= \sqrt{n}\, u_{ij} \quad (i \neq j), \\ b_{ii} &= n\rho'_i + \sqrt{n}\, v_{ii}, & b_{ig} &= \sqrt{n}\, v_{ig} \quad (i \neq g), \\ c_{gg} &= n + \sqrt{n}\, w_{gg}, & c_{gh} &= \sqrt{n}\, w_{gh} \quad (g \neq h), \end{aligned}\right\} \qquad \dots\dots(10)$$

$$(i, j = 1, \dots, p;\ g, h = 1, \dots, q).$$

The distribution of the u's, v's and w's approach that of $\frac{1}{2}(p+q)(p+q+1)$ normal variates whose means are zero and whose second moments are specified in the following statements:

(i) *any v or w which has at least one suffix number $> p$ is uncorrelated with all the others*;

(ii) *any member of one of the sets (u_{ii}, v_{ii}, w_{ii}), $(i = 1, \dots, p)$, $(u_{ij}, v_{ij}, v_{ji}, w_{ij})$, $(i, j = 1, \dots, p;\ i \neq j)$ is uncorrelated with all the members of all the other sets*;

(iii) *for $i, j = 1, \dots, p$ we have*

$$\left.\begin{aligned} \mathscr{E}(u_{ii}^2) = \mathscr{E}(w_{ii}^2) &= 2, \\ \mathscr{E}(v_{ii}^2) &= 1 + \rho_i'^2, \\ \mathscr{E}(u_{ii} w_{ii}) &= 2\rho_i'^2, \\ \mathscr{E}(u_{ii} v_{ii}) = \mathscr{E}(w_{ii} v_{ii}) &= 2\rho'_i, \\ \mathscr{E}(u_{ij}^2) = \mathscr{E}(v_{ij}^2) = \mathscr{E}(w_{ij}^2) &= 1 \quad (i \neq j), \\ \mathscr{E}(u_{ij} w_{ij}) = \mathscr{E}(v_{ij} v_{ji}) &= \rho'_i \rho'_j \quad (i \neq j), \\ \mathscr{E}(u_{ij} v_{ij}) = \mathscr{E}(w_{ij} v_{ij}) &= \rho'_j \quad (i \neq j). \end{aligned}\right\} \qquad \dots\dots(11)$$

This follows immediately from the well-known central limit theorem.* The second moments may easily be computed by virtue of (3).

COROLLARY. *Under the assumption (6) any u, v or w which has at least one suffix number $> s$ is uncorrelated with all the others, and $\mathscr{E}(v_{ii}^2) = 1$ for $i = s+1, \ldots, p$.*

This results from (11) on putting $\rho'_i = 0$ for $i = s+1, \ldots, p$.

3. We may now find the limiting distribution of $r_{s+1}, \ldots, r_p$, the $p-s$ smallest correlations.

We substitute (10) in (4) and then divide each element by n. There results the equation

$$\begin{vmatrix}
r\left(1+\frac{u_{11}}{\sqrt{n}}\right) & \cdots & \frac{ru_{1s}}{\sqrt{n}} & \frac{ru_{1,s+1}}{\sqrt{n}} & \cdots & \frac{ru_{1p}}{\sqrt{n}} & \rho'_1+\frac{v_{11}}{\sqrt{n}} & \cdots & \frac{v_{1s}}{\sqrt{n}} & \frac{v_{1,s+1}}{\sqrt{n}} & \cdots & \frac{v_{1q}}{\sqrt{n}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{ru_{s1}}{\sqrt{n}} & \cdots & r\left(1+\frac{u_{ss}}{\sqrt{n}}\right) & \frac{ru_{s,s+1}}{\sqrt{n}} & \cdots & \frac{ru_{sp}}{\sqrt{n}} & \frac{v_{s1}}{\sqrt{n}} & \cdots & \rho'_s+\frac{v_{ss}}{\sqrt{n}} & \frac{v_{s,s+1}}{\sqrt{n}} & \cdots & \frac{v_{sq}}{\sqrt{n}} \\
\frac{ru_{s+1,1}}{\sqrt{n}} & \cdots & \frac{ru_{s+1,s}}{\sqrt{n}} & r\left(1+\frac{u_{s+1,s+1}}{\sqrt{n}}\right) & \cdots & \frac{ru_{s+1,p}}{\sqrt{n}} & \frac{v_{s+1,1}}{\sqrt{n}} & \cdots & \frac{v_{s+1,s}}{\sqrt{n}} & \frac{v_{s+1,s+1}}{\sqrt{n}} & \cdots & \frac{v_{s+1,q}}{\sqrt{n}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{ru_{p1}}{\sqrt{n}} & \cdots & \frac{ru_{ps}}{\sqrt{n}} & \frac{ru_{p,s+1}}{\sqrt{n}} & \cdots & r\left(1+\frac{u_{pp}}{\sqrt{n}}\right) & \frac{v_{p1}}{\sqrt{n}} & \cdots & \frac{v_{ps}}{\sqrt{n}} & \frac{v_{p,s+1}}{\sqrt{n}} & \cdots & \frac{v_{pq}}{\sqrt{n}} \\
\rho'_1+\frac{v_{11}}{\sqrt{n}} & \cdots & \frac{v_{s1}}{\sqrt{n}} & \frac{v_{s+1,1}}{\sqrt{n}} & \cdots & \frac{v_{p1}}{\sqrt{n}} & r\left(1+\frac{w_{11}}{\sqrt{n}}\right) & \cdots & \frac{rw_{1s}}{\sqrt{n}} & \frac{rw_{1,s+1}}{\sqrt{n}} & \cdots & \frac{rw_{1q}}{\sqrt{n}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{v_{1s}}{\sqrt{n}} & \cdots & \rho'_s+\frac{v_{ss}}{\sqrt{n}} & \frac{v_{s+1,s}}{\sqrt{n}} & \cdots & \frac{v_{ps}}{\sqrt{n}} & \frac{rw_{s1}}{\sqrt{n}} & \cdots & r\left(1+\frac{w_{ss}}{\sqrt{n}}\right) & \frac{rw_{s,s+1}}{\sqrt{n}} & \cdots & \frac{rw_{sq}}{\sqrt{n}} \\
\frac{v_{1,s+1}}{\sqrt{n}} & \cdots & \frac{v_{s,s+1}}{\sqrt{n}} & \frac{v_{s+1,s+1}}{\sqrt{n}} & \cdots & \frac{v_{p,s+1}}{\sqrt{n}} & \frac{rw_{s+1,1}}{\sqrt{n}} & \cdots & \frac{rw_{s+1,s}}{\sqrt{n}} & r\left(1+\frac{w_{s+1,s+1}}{\sqrt{n}}\right) & \cdots & \frac{rw_{s+1,q}}{\sqrt{n}} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{v_{1q}}{\sqrt{n}} & \cdots & \frac{v_{sq}}{\sqrt{n}} & \frac{v_{s+1,q}}{\sqrt{n}} & \cdots & \frac{v_{pq}}{\sqrt{n}} & \frac{rw_{q1}}{\sqrt{n}} & \cdots & \frac{rw_{qs}}{\sqrt{n}} & \frac{rw_{q,s+1}}{\sqrt{n}} & \cdots & r\left(1+\frac{w_{qq}}{\sqrt{n}}\right)
\end{vmatrix} = 0. \qquad \ldots\ldots(12)$$

The equation (12) has $p-s$ roots which are $o(1)$ for large n. To evaluate these, we substitute $n^{-\frac{1}{2}}\eta$ for r in (12), delete the common factor $n^{-\frac{1}{2}}$ from the rows

$$s+1, s+2, \ldots, p, p+s+1, p+s+2, \ldots, p+q,$$

* Cf. Cramér (1937), pp. 113–14.

and then let $n \to \infty$. There results the equation

$$\begin{vmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 & \rho_1' & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & \rho_s' & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -\eta & \cdots & 0 & v_{s+1,1} & \cdots & v_{s+1,s} & v_{s+1,s+1} & \cdots & v_{s+1,q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 0 & \cdots & -\eta & v_{p1} & \cdots & v_{ps} & v_{p,s+1} & \cdots & v_{pq} \\ \rho_1' & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \rho_s' & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ v_{1,s+1} & \cdots & v_{s,s+1} & v_{s+1,s+1} & \cdots & v_{p,s+1} & 0 & \cdots & 0 & -\eta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{1q} & \cdots & v_{sq} & v_{s+1,q} & \cdots & v_{pq} & 0 & \cdots & 0 & 0 & \cdots & -\eta \end{vmatrix} = 0,$$

i.e.

$$\begin{vmatrix} -\eta & \cdots & 0 & v_{s+1,s+1} & \cdots & v_{s+1,q} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & -\eta & v_{p,s+1} & \cdots & v_{pq} \\ v_{s+1,s+1} & \cdots & v_{p,s+1} & -\eta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ v_{s+1,q} & \cdots & v_{pq} & 0 & \cdots & -\eta \end{vmatrix} = 0. \qquad \ldots\ldots(13)$$

By (9) the left-hand side of (13) is equal to

$$(-\eta)^{q-p} \begin{vmatrix} d_{s+1,s+1} - \eta^2 & \cdots & d_{s+1,q} \\ \cdots & \cdots & \cdots \\ d_{q,s+1} & \cdots & d_{qq} - \eta^2 \end{vmatrix}, \qquad \ldots\ldots(14)$$

where

$$d_{ij} = \sum_{g=s+1}^{q} v_{ig} v_{jg} \quad (i, j = s+1, \ldots, p). \qquad \ldots\ldots(15)$$

Let $\zeta_{s+1}', \ldots, \zeta_p'$, in descending order of magnitude, be the latent roots of the matrix $\| d_{ij} \|$. Then the $p-s$ roots of (12) which are $o(1)$ for large n are

$$n^{-\frac{1}{2}} \zeta_i'^{\frac{1}{2}} + o(n^{-\frac{1}{2}}) \quad (i = s+1, \ldots, p).$$

If we define $\zeta_{s+1}, \ldots, \zeta_p$ by putting

$$r_i = n^{-\frac{1}{2}} \zeta_i^{\frac{1}{2}} \quad (i = s+1, \ldots, p), \qquad \ldots\ldots(16)$$

then the ζ's have the same limiting distribution as the ζ''s. Hence the limiting distribution of the ζ_i may be derived as the distribution of the latent roots of $\| d_{ij} \|$, in which the v's are regarded as having a distribution which is the limiting distribution described in Lemma 2. By virtue of the Corollary this is the distribution of $(p-s)(q-s)$ mutually independent normal variates with zero mean

and unit standard deviation. Therefore* the limiting distribution of the $\zeta_i = nr_i^2$ has the density function

$$2^{-\frac{1}{2}(p-s)(q-s)}\pi^{\frac{1}{2}(p-s)}\left\{\prod_{i=1}^{p-s}\Gamma\tfrac{1}{2}(q-s-i+1)\Gamma\tfrac{1}{2}i\right\}^{-1}\left\{\prod_{i=s+1}^{p}\prod_{j=i+1}^{p}(\zeta_i-\zeta_j)\right\}$$
$$\times\left(\prod_{i=s+1}^{p}\zeta_i\right)^{\frac{1}{2}(q-p-1)}\exp\left(-\tfrac{1}{2}\sum_{i=s+1}^{p}\zeta_i\right), \quad \ldots\ldots(17)$$

$$\infty > \zeta_{s+1} \geqslant \ldots \geqslant \zeta_p \geqslant 0.$$

The transformation $\zeta_i = \eta_i^2$ gives the following density for the limiting distribution of the $\eta_i = n^{\frac{1}{2}} r_i$ $(i=s+1, \ldots, p)$:

$$f_1(\eta_{s+1}, \ldots, \eta_p) = 2^{p-s-\frac{1}{2}(p-s)(q-s)}\pi^{\frac{1}{2}(p-s)}\left\{\prod_{i=1}^{p-s}\Gamma\tfrac{1}{2}(q-s-i+1)\,\Gamma\tfrac{1}{2}i\right\}^{-1}$$
$$\times\left\{\prod_{i=s+1}^{p}\prod_{j=i+1}^{p}(\eta_i^2-\eta_j^2)\right\}\left(\prod_{i=s+1}^{p}\eta_i\right)^{q-p}\exp\left(-\tfrac{1}{2}\sum_{i=s+1}^{p}\eta_i^2\right), \quad \ldots\ldots(18)$$

$$\infty > \eta_{s+1} \geqslant \ldots \geqslant \eta_p \geqslant 0.$$

4. We now proceed to find the limiting distribution of $r_1, \ldots, r_s$. By virtue of (9) we may write the left-hand side of (4) as

$$r^{q-p}\,|\,\mathbf{C}\,|\,.\,|\,r^2\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\,|,$$

where

$$\mathbf{A} = \begin{Vmatrix} a_{11} & \cdots & a_{1p} \\ \cdots & \cdots & \cdots \\ a_{p1} & \cdots & a_{pp} \end{Vmatrix}, \quad \mathbf{B} = \begin{Vmatrix} b_{11} & \cdots & b_{1q} \\ \cdots & \cdots & \cdots \\ b_{p1} & \cdots & b_{pq} \end{Vmatrix}, \quad \mathbf{C} = \begin{Vmatrix} c_{11} & \cdots & c_{1q} \\ \cdots & \cdots & \cdots \\ c_{q1} & \cdots & c_{qq} \end{Vmatrix}. \quad \ldots\ldots(19)$$

Hence, if we set $\theta_i = r_i^2 \quad (i=1, \ldots, p),$ ......(20)

the θ_i are the roots of the equation

$$|\,\mathbf{B}\mathbf{C}^{-1}\mathbf{B}'-\theta\mathbf{A}\,| = 0. \quad \ldots\ldots(21)$$

Substituting (10) in (21) and dividing each element by n, we get

$$|\,(\boldsymbol{\Delta}+n^{-\frac{1}{2}}\mathbf{V})(\mathbf{I}+n^{-\frac{1}{2}}\mathbf{W})^{-1}(\boldsymbol{\Delta}'+n^{-\frac{1}{2}}\mathbf{V}')-\theta(\mathbf{I}+n^{-\frac{1}{2}}\mathbf{U})\,| = 0, \quad \ldots\ldots(22)$$

where

$$\mathbf{U} = \begin{Vmatrix} u_{11} & \cdots & u_{1p} \\ \cdots & \cdots & \cdots \\ u_{p1} & \cdots & u_{pp} \end{Vmatrix}, \quad \mathbf{V} = \begin{Vmatrix} v_{11} & \cdots & v_{1q} \\ \cdots & \cdots & \cdots \\ v_{p1} & \cdots & v_{pq} \end{Vmatrix}, \quad \mathbf{W} = \begin{Vmatrix} w_{11} & \cdots & w_{1q} \\ \cdots & \cdots & \cdots \\ w_{q1} & \cdots & w_{qq} \end{Vmatrix}, \quad \ldots\ldots(23)$$

$$\boldsymbol{\Delta} = \begin{Vmatrix} \rho_1' & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \rho_p' & 0 & \cdots & 0 \end{Vmatrix} \quad \ldots\ldots(24)$$

* Hsu (1939), pp. 256–7.

Neglecting higher powers of $n^{-\frac{1}{2}}$ we write $\mathbf{I}-n^{-\frac{1}{2}}\mathbf{W}$ for $(\mathbf{I}+n^{-\frac{1}{2}}\mathbf{W})^{-1}$ and carry out the matrix multiplication to the term with $n^{-\frac{1}{2}}$ as a factor. There results the equation

$$|\boldsymbol{\Delta}\boldsymbol{\Delta}'+n^{-\frac{1}{2}}(\mathbf{V}\boldsymbol{\Delta}'+\boldsymbol{\Delta}\mathbf{V}'-\boldsymbol{\Delta}\mathbf{W}\boldsymbol{\Delta}')-\theta(\mathbf{I}+n^{-\frac{1}{2}}\mathbf{U})| = 0, \qquad \text{......(25)}$$

i.e.

$$\begin{vmatrix} \rho_1'^2-\theta+n^{-\frac{1}{2}}(2\rho_1'v_{11}-\rho_1'^2w_{11}-\theta u_{11}) & n^{-\frac{1}{2}}(\rho_1'v_{21}+\rho_2'v_{12}-\rho_1'\rho_2'w_{12}-\theta u_{12}) & \cdots & n^{-\frac{1}{2}}(\rho_1'v_{p1}+\rho_p'v_{1p}-\rho_1'\rho_p'w_{1p}-\theta u_{1p}) \\ n^{-\frac{1}{2}}(\rho_1'v_{21}+\rho_2'v_{12}-\rho_1'\rho_2'w_{12}-\theta u_{12}) & \rho_2'^2-\theta+n^{-\frac{1}{2}}(2\rho_2'v_{22}-\rho_2'^2w_{22}-\theta u_{22}) & \cdots & n^{-\frac{1}{2}}(\rho_2'v_{p2}+\rho_p'v_{2p}-\rho_2'\rho_p'w_{2p}-\theta u_{2p}) \\ \cdots & \cdots & \cdots & \cdots \\ n^{-\frac{1}{2}}(\rho_1'v_{p1}+\rho_p'v_{1p}-\rho_1'\rho_p'w_{1p}-\theta u_{1p}) & n^{-\frac{1}{2}}(\rho_2'v_{p2}+\rho_p'v_{2p}-\rho_2'\rho_p'w_{2p}-\theta u_{2p}) & \cdots & \rho_p'^2-\theta+n^{-\frac{1}{2}}(2\rho_p'v_{pp}-\rho_p'^2w_{pp}-\theta u_{pp}) \end{vmatrix} = 0. \qquad \text{......(26)}$$

On account of (5) there are μ_1 roots of (26) which are $\rho_1^2+o(1)$ for large n. To evaluate these we substitute $\rho_1^2+n^{-\frac{1}{2}}\rho_1\zeta$ for θ in (26). Since the first μ_1 of the ρ''s are equal to ρ_1, there will be a common factor $n^{-\frac{1}{2}}$ in each of the first μ_1 rows. After deleting this factor and then letting $n\to\infty$, we get the equation

$$\begin{vmatrix} 2v_{11}-\rho_1(u_{11}+w_{11})-\zeta & v_{12}+v_{21}-\rho_1(u_{12}+w_{12}) & \cdots & v_{1\mu_1}+v_{\mu_1 1}-\rho_1(u_{1\mu_1}+w_{1\mu_1}) \\ v_{12}+v_{21}-\rho_1(u_{12}+w_{12}) & 2v_{22}-\rho_1(u_{22}+w_{22})-\zeta & \cdots & v_{2\mu_1}+v_{\mu_1 2}-\rho_1(u_{2\mu_1}+w_{2\mu_1}) \\ \cdots & \cdots & \cdots & \cdots \\ v_{1\mu_1}+v_{\mu_1 1}-\rho_1(u_{1\mu_1}+w_{1\mu_1}) & v_{2\mu_1}+v_{\mu_1 2}-\rho_1(u_{2\mu_1}+w_{2\mu_1}) & \cdots & 2v_{\mu_1\mu_1}-\rho_1(u_{\mu_1\mu_1}+w_{\mu_1\mu_1})-\zeta \end{vmatrix} = 0, \qquad \text{......(27)}$$

i.e.

$$\begin{vmatrix} z_{11}-\zeta & \cdots & z_{1\mu_1} \\ \cdots & \cdots & \cdots \\ z_{\mu_1 1} & \cdots & z_{\mu_1\mu_1}-\zeta \end{vmatrix} = 0, \qquad \text{......(28)}$$

where

$$z_{ij} = v_{ij}+v_{ji}-\rho_1(u_{ij}+w_{ij}) \quad (i,j=1,\ldots,\mu_1). \qquad \text{......(29)}$$

Let $\zeta_1',\ldots,\zeta_{\mu_1}'$ be the roots, in descending order of magnitude, of (28). Then the μ_1 roots of (26) which are $\rho_1^2+o(1)$ for large n are

$$\rho_1^2+n^{-\frac{1}{2}}\rho_1\zeta_i'+o(n^{-\frac{1}{2}}) \quad (i=1,\ldots,\mu_1).$$

If we define $\zeta_1,\ldots,\zeta_{\mu_1}$ by putting

$$\theta_i = \rho_1^2+n^{-\frac{1}{2}}\rho_1\zeta_i \quad (i=1,\ldots,\mu_1), \qquad \text{......(30)}$$

then the ζ's have the same limiting distribution as the ζ''s. Hence the limiting distribution of the ζ_i may be derived as the distribution of the latent roots of

the matrix $\|z_{ij}\|$, in which the u's, v's and w's are regarded as having a distribution which is the limiting distribution described in Lemma 2.

Now, if all the u's, v's and w's are normal variates with zero mean, so are the z's. Owing to the fact (ii) of Lemma 2 all the z's are uncorrelated. Using the formulae (11) to calculate the variances of the z's, we easily obtain

$$\mathscr{E}(z_{ii}^2) = 4(1-\rho_1^2)^2, \quad \mathscr{E}(z_{ij}^2) = 2(1-\rho_1^2)^2 \quad (i \neq j), \qquad \text{......(31)}$$
$$(i, j = 1, \ldots, \mu_1).$$

Hence
$$z_{ii} = 2(1-\rho_1^2)\,t_{ii}, \quad z_{ij} = \sqrt{2}\,(1-\rho_1^2)\,t_{ij} \quad (i \neq j), \qquad \text{......(32)}$$
$$(i, j = 1, \ldots, \mu_1),$$

where the t's are mutually independent normal variates with zero mean and unit standard deviation.

Setting $\zeta = 2(1-\rho_1^2)\eta$ in (28), we get, by (32),

$$\begin{vmatrix} t_{11}-\eta & 2^{-\frac{1}{2}}t_{12} & \ldots & 2^{-\frac{1}{2}}t_{1\mu_1} \\ 2^{-\frac{1}{2}}t_{21} & t_{22}-\eta & \ldots & 2^{-\frac{1}{2}}t_{2\mu_1} \\ \cdots & \cdots & \cdots & \cdots \\ 2^{-\frac{1}{2}}t_{\mu_1 1} & 2^{-\frac{1}{2}}t_{\mu_1 2} & \ldots & t_{\mu_1\mu_1}-\eta \end{vmatrix} = 0. \qquad \text{......(33)}$$

Let $\eta_1', \ldots, \eta_{\mu_1}'$ be the roots of (33) in descending order of magnitude.

The density function

$$(2\pi)^{-\frac{1}{4}\mu_1(\mu_1+1)}\exp\{-\tfrac{1}{2}(t_{11}^2+\ldots+t_{\mu_1\mu_1}^2+t_{12}^2+\ldots+t_{\mu_1-1,\mu_1}^2)\} \qquad \text{......(34)}$$

is equal to
$$(2\pi)^{-\frac{1}{4}\mu_1(\mu_1+1)}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{\mu_1}\eta_i'^2\right), \qquad \text{......(35)}$$

which is a function of the latent roots only. Hence* the distribution of the latent roots has the density

$$f(\eta_1', \ldots, \eta_{\mu_1}') = 2^{-\frac{1}{4}\mu_1(\mu_1+1)}\left(\prod_{i=1}^{\mu_1}\Gamma\tfrac{1}{2}i\right)^{-1}\left\{\prod_{i=1}^{\mu_1}\prod_{j=i+1}^{\mu_1}(\eta_i'-\eta_j')\right\}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{\mu_1}\eta_i'^2\right), \quad \text{......(36)}$$
$$\infty > \eta_1' \geqslant \ldots \geqslant \eta_{\mu_1}' > -\infty.$$

The density function (36) represents the limiting distribution of $\eta_1', \ldots, \eta_{\mu_1}'$, where

$$\theta_i = \rho_1^2 + 2n^{-\frac{1}{2}}\rho_1(1-\rho_1^2)\,\eta_i' \quad (i = 1, \ldots, \mu_1). \qquad \text{......(37)}$$

Hence
$$r_i = \theta_i^{\frac{1}{2}} = \rho_1 + n^{-\frac{1}{2}}(1-\rho_1^2)\,\eta_i' + o(n^{-\frac{1}{2}}) \quad (i = 1, \ldots, \mu_1).$$

If we define $\eta_1, \ldots, \eta_{\mu_1}$ by

$$r_i = \rho_1 + n^{-\frac{1}{2}}(1-\rho_1^2)\,\eta_i \quad (i = 1, \ldots, \mu_1), \qquad \text{......(38)}$$

then the $\eta_i = n^{\frac{1}{2}}(1-\rho_1^2)^{-1}(r_i-\rho_1)$ have the same limiting distribution as the η_i'. Hence the limiting distribution of $\eta_1, \ldots, \eta_{\mu_1}$ has the density $f(\eta_1, \ldots, \eta_{\mu_1})$.

In exactly the same manner we may prove that for $k = 1, \ldots, \nu$ the

$$\eta_i = n^{\frac{1}{2}}(1-\rho_k^2)^{-1}(r_i-\rho_k) \quad (i = \mu_1+\ldots+\mu_{k-1}+1, \ldots, \mu_1+\ldots+\mu_k)$$

* Hsu (1939), p. 256, Theorem 2.

have the limiting distribution represented by the density

$$f(\eta_{\mu_1+\ldots+\mu_{k-1}+1}, \ldots, \eta_{\mu_1+\ldots+\mu_k}),$$

where, in general,

$$f(x_1, \ldots, x_m) = 2^{-\frac{1}{4}m(m+1)}\left(\prod_{i=1}^{m} \Gamma\tfrac{1}{2}i\right)^{-1}\left\{\prod_{i=1}^{m}\prod_{j=i+1}^{m}(x_i-x_j)\right\}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{m} x_i^2\right), \quad \ldots\ldots(39)$$

$$\infty > x_1 \geqslant \ldots \geqslant x_m > -\infty.$$

Furthermore, the sets $(\eta_1, \ldots, \eta_{\mu_1})$, $(\eta_{\mu_1+1}, \ldots, \eta_{\mu_1+\mu_2})$, $\ldots$, $(\eta_{\mu_1+\ldots+\mu_{\nu-1}+1}, \ldots, \eta_s)$ are such that the equations corresponding to (27) for two different sets involve only mutually uncorrelated u's, v's and w's owing to (ii) of Lemma 2. Therefore the limiting distribution must be such that these sets are independent of one another. Also, recalling (14) and (i) of Lemma 2, it is seen that the limiting distribution of $\eta_1, \ldots, \eta_p$ must be such that the sets $(\eta_1, \ldots, \eta_{\mu_1})$, $(\eta_{\mu_1+1}, \ldots, \eta_{\mu_1+\mu_2})$, $\ldots$, $(\eta_{\mu_1+\ldots+\mu_{\nu-1}+1}, \ldots, \eta_s)$ and $(\eta_{s+1}, \ldots, \eta_p)$ are independent of one another.

In conclusion we sum up the results in the following theorem:

THEOREM. *Let the population canonical correlations be $\rho'_1, \ldots, \rho'_p$, where*

$$\rho'_1 = \ldots = \rho'_{\mu_1} = \rho_1,$$
$$\rho'_{\mu_1+1} = \ldots = \rho'_{\mu_1+\mu_2} = \rho_2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\rho'_{\mu_1+\ldots+\mu_{\nu-1}+1} = \ldots = \rho'_s = \rho_\nu,$$
$$\rho'_{s+1} = \ldots = \rho'_p = 0,$$
$$\rho_1 > \rho_2 > \ldots > \rho_\nu > 0.$$

Let the sample canonical correlations be $r_1, \ldots, r_p$, where

$$r_1 \geqslant r_2 \geqslant \ldots \geqslant r_p.$$

Let

$$\eta_i = n^{\frac{1}{2}}(1-\rho_i'^2)^{-1}(r_i-\rho'_i) \quad (i=1, \ldots, p).$$

Then the limiting distribution of $\eta_1, \ldots, \eta_p$ is represented by the density function

$$f(\eta_1, \ldots, \eta_{\mu_1})f(\eta_{\mu_1+1}, \ldots, \eta_{\mu_1+\mu_2})\ldots f(\eta_{\mu_1+\ldots+\mu_{\nu-1}+1}, \ldots, \eta_s)f_1(\eta_{s+1}, \ldots, \eta_p),$$

where the functions f and f_1 are given by (39) *and* (18) *respectively.*

REFERENCES

CRAMÉR, H. (1937). *Random Variables and Probability Distributions.* Camb. Univ. Press.

GIRSCHICK, M. A. (1939). "On the sampling theory of roots of determinantal equations." *Ann. Math. Statist.* **10**, 205.

HOTELLING, H. (1936). "Relations between two sets of variates." *Biometrika*, **28**, 321–77

HSU, P. L. (1939). "On the distribution of roots of certain determinantal equations." *Ann. Eugen., Lond.*, **9**, 250–8.

—— (1940). "On the limiting distribution of roots of a determinantal equation." *Proc. Lond. Math. Soc.* (at press).

MADOW, W. G. (1938). "Contributions to the theory of multi-variate statistical analysis." *Trans. Amer. Math. Soc.* **44**, 454.

THE APPLICATION OF MAXIMUM LIKELIHOOD TO DOSAGE-MORTALITY CURVES

By F. GARWOOD, Ph.D.

1. Introduction

Many papers have been written on the fitting of dosage-mortality curves; in particular, a paper by Irwin & Cheeseman (1939) summarizes the methods which have been adopted hitherto. It is felt, however, that there is some mathematical interest in the subject which is worth emphasizing.

A typical problem occurs when studying the effect of some drug on a particular kind of animal. It is assumed that there is a population of animals, and associated with each individual animal is a certain lethal dose of the drug, such that the animal would always be killed by a stronger dose and would survive a weaker one. There is independent biological evidence for assuming that the logarithms of the lethal doses are normally distributed throughout the population, so that if the proportion of animals expected to survive a given dose is converted into a probit (i.e. an equivalent normal deviate + 5), then the above assumption is equivalent to stating that the probits are linearly related to the logs of the doses. If the mean (or median) log lethal dose is m and the standard deviation is σ, then the linear relation between probit and log lethal dose is

$$Y = \alpha + \beta x,$$

where
$$\sigma = \frac{1}{\beta} \quad \text{and} \quad m = \frac{5-\alpha}{\beta}.$$

The experimental material consists of k groups, drawn at random from the population, of $n_1, n_2, \ldots, n_k$ animals, which are given doses with logs $x_1, x_2, \ldots, x_k$, from which there are $s_1, s_2, \ldots, s_k$ survivors, and $n_1 - s_1, n_2 - s_2, \ldots, n_k - s_k$ deaths.

The treatment which has hitherto been applied to data of this kind consists of obtaining from tables the probits $y_1, y_2, \ldots$ corresponding to the proportions of survivors $q_1 = s_1/n_1$, $q_2 = s_2/n_2, \ldots$, plotting the y's against the corresponding log doses $x_1, x_2, \ldots$ and fitting a line to the points, bearing in mind the following considerations.

Since $dq/dy = -Z$, Z being the ordinate of the normal curve, and as the variance of q is PQ/n, $Q\ (= 1-P)$ being the expected proportion of survivors, it follows that the variance of y is PQ/nZ^2 which in general varies along the line. Thus different weights must be used for the various probits in fitting the line. The effects of using different methods of calculating the weighting coefficient $w = nZ^2/PQ$ (reciprocal of the variance) have been compared by Irwin & Cheeseman (1939).

A further difficulty occurred in the cases of zero and all survivors, for which the corresponding probits are infinite. Fisher's method (Bliss, 1935), using the method of maximum likelihood, overcame this difficulty by replacing the infinite probit by a working or fictitious probit in the regression equations.

Mathematically, Fisher's exact method of calculation is as follows (see also Bliss, 1938 and Fisher & Yates, 1938). Assume rough values a_1 and b_1 in the relation

$$Y = a_1 + b_1 x.$$

For each value of x this formula gives P and Q (the areas of the normal curve up to and beyond the probit Y), Z (the ordinate at Y) and $w = nZ^2/PQ$. The regression is then found between the variate

$$y = Y + \eta = a_1 + b_1 x + \eta, \qquad \ldots\ldots(1)$$

where

$$\eta = \frac{Q-q}{Z},$$

and x, giving weights w to the former. The result is a new regression equation

$$Y = a_2 + b_2 x,$$

where

$$b_2 = \frac{Swx(y-\bar{y})}{Sw(x-\bar{x})^2},$$

and

$$a_2 = \bar{y} - b_2 \bar{x}.$$

It is to be noted that this form of the regression equation is more convenient for our purposes than the form $Y = \bar{y} + b_2(x-\bar{x})$. Substituting the values of y from (1), it follows that

$$b_2 = b_1 + \frac{Swx(\eta-\bar{\eta})}{Sw(x-\bar{x})^2}, \quad \text{and} \quad a_2 = a_1 + \bar{\eta} - \bar{x}(b_2 - b_1).$$

Hence the changes $\delta a = a_2 - a_1$ and $\delta b = b_2 - b_1$ in the regression coefficients of $y = Y + \eta$ on x are in fact the regression coefficients of η on x, and they can be regarded as the solutions of the normal equations

$$\delta a Sw + \delta b Swx = Sw\eta, \qquad \ldots\ldots(2)$$

$$\delta a Swx + \delta b Swx^2 = Swx\eta. \qquad \ldots\ldots(3)$$

The new regression equation $Y = a_2 + b_2 x$ is then made the basis of a similar calculation; i.e. the regression is calculated between $a_2 + b_2 x + \eta$ and x (the values of η will be changed since Q and Z are in general altered), giving another equation $Y = a_3 + b_3 x$ and so on. The process is continued until no change occurs in the coefficients.

It is possible that some arithmetical labour might be saved by obtaining the corrections $\delta a, \delta b$ to the regression coefficients a, b at each stage, instead of the new coefficients $a + \delta a, b + \delta b$, by calculating the regression between η and x, but this has not been investigated. The process of obtaining the corrections is illustrated later (Table III).

In general the successive coefficients a_n and b_n will converge to limits (and the successive corrections $\delta a, \delta b$ to zero), and it is not difficult to see that these limits are the solutions of the maximum likelihood equations. The process is, in fact, the same as the general method suggested by Fisher for solving maximum likelihood equations, as will be shown later. Also, a consideration of the foundations of this method shows that there is a method of obtaining the maximum likelihood estimates which is slightly more rapid (as regards numbers of approximations) than that outlined above.

It is one of the objects of this note to point out that the problem may be regarded as one of estimating the parameters m and σ (or equally, α and β), i.e. of fitting a normal curve to the data. The fact that this is equivalent to fitting a line to the theoretical relation between probits and log lethal doses is only a consequence of the special nature of the normal distribution. It may happen in other applications that the distribution of the log lethal doses is not normal and cannot be normalized by any transformation of the log lethal doses but has another form depending on unknown parameters; then the problem can only be regarded in general as the estimation of parameters from observations.

In the case of the normal distribution the method of plotting probits is of course a very convenient method of representing the data and of obtaining a good general picture, but it is to be emphasized that from the theoretical viewpoint it is at least equally important to interpret the problem as one of estimating parameters as to regard it as a problem of fitting a regression line in the ordinary sense of the term.

2. General maximum likelihood estimates

It will be convenient to recapitulate the method used by Fisher for solving maximum likelihood equations (see, e.g., Koshal, 1933). A sample $x_1, x_2, \ldots, x_n$ is drawn at random from a population of which the frequency function has a known form depending on s unknown parameters $\theta_1, \theta_2, \ldots, \theta_s$ so that the probability of obtaining the sample is

$$P(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_s).$$

If the variates x are independent of each other, as in the case of successive samples from the same population, then P is the product of n probability functions $f(x_1; \theta_1, \theta_2, \ldots), f(x_2; \theta_1, \theta_2, \ldots), \ldots$. On the other hand, they will not be independent if they are a set of frequencies with a fixed total.

The maximum likelihood estimates of the unknown parameters $\theta_1, \theta_2, \ldots$ based on the information provided by the sample, are the values which satisfy the maximum likelihood equations $\partial P/\partial\theta_1 = 0$, $\partial P/\partial\theta_2 = 0, \ldots$. Using the likelihood function

$$L = L(x_1, x_2, \ldots; \theta_1, \theta_2, \ldots) = \log P,$$

the estimates must be solutions of $\partial L/\partial\theta_1 = 0$, etc.

Since they are functions of the sample, these solutions can be written $T_1 = T_1(x_1, x_2, \ldots), T_2, \ldots$, etc.; if approximations $t_1, t_2, \ldots$ to $T_1, T_2, \ldots$ are found by some rough method, suppose that $T_1 = t_1 + \delta t_1$, etc., so that $\delta t_1, \delta t_2, \ldots$ are the errors in the approximations. Then we have

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \theta_1}(T_1, T_2, \ldots) \\ &= \frac{\partial L}{\partial \theta_1}(t_1, t_2, \ldots) + \delta t_1 \frac{\partial^2 L}{\partial \theta_1^2}(t_1, t_2, \ldots) + \delta t_2 \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}(t_1, t_2, \ldots) + \ldots, \end{aligned} \qquad \ldots\ldots(4)$$

etc., ignoring terms in δt_1^2, etc. Writing

$$\frac{\partial L}{\partial \theta_1}(t_1, t_2, \ldots) = L_1,$$

$$\frac{\partial^2 L}{\partial \theta_1^2}(t_1, t_2, \ldots) = L_{11}, \text{ etc.,}$$

the equations become $\quad L_{11}\,\delta t_1 + L_{12}\,\delta t_2 + \ldots = -L_1$, etc.

or

$$\sum_{j=1}^{s} L_{ij}\,\delta t_j = -L_i, \quad (i = 1, \ldots, s). \qquad \ldots\ldots(5)$$

The solutions of these equations can be written

$$\delta t_i = -\sum_{j=1}^{s} l_{ij} L_j, \quad (i = 1, \ldots, s),$$

where, in matrix notation, $\{l_{ij}\} = \{L_{ij}\}^{-1}$, the reciprocal matrix of $\{L_{ij}\}$. Thus if Δ is the determinant formed by the elements L_{ij}, and Δ_{ij} the determinant obtained by omitting row i and column j, then

$$l_{11} = \frac{\Delta_{11}}{\Delta}, \quad l_{12} = -\frac{\Delta_{12}}{\Delta}, \quad l_{13} = \frac{\Delta_{13}}{\Delta}, \quad \ldots, \quad l_{ij} = \frac{(-)^{i+j}\Delta_{ij}}{\Delta}.$$

For two variables

$$l_{11} = -\frac{L_{22}}{\Delta}, \quad l_{12} = l_{21} = \frac{L_{12}}{\Delta}, \quad l_{22} = -\frac{L_{11}}{\Delta}, \quad \Delta = L_{11}L_{22} - L_{12}^2,$$

and $\quad \delta t_1 = -(l_{11}L_1 + l_{12}L_2), \quad \delta t_2 = -(l_{21}L_1 + l_{22}L_2).$

The corrections δt will not be exact, since terms of higher order have been omitted from (4); however, if the process is repeated with $t_1 + \delta t_1, t_2 + \delta t_2, \ldots$, now used as the first approximations, the corrections obtained will be of the next order of smallness, and the process can be continued until the approximations t are as near as desired to the exact estimates T.

The coefficients L_{ij} are functions of the observations as well as the approximations t; it is in some practical cases more convenient to replace the x's by their expected values, i.e. the values which they would be expected to have if $\theta_1 = t_1$,

$\theta_2 = t_2 \ldots$ Let Λ_{ij} be the resulting values of L_{ij}, so that equations (5) become

$$\sum_{j=1}^{s} \Lambda_{ij}\, \delta t_j = -L_i, \quad (i = 1, \ldots, s)$$

with solutions

$$\delta t_i = -\sum_{j=1}^{s} \lambda_{ij} L_j, \quad (i = 1, \ldots, s)$$

where

$$\{\lambda_{ij}\} = \{\Lambda_{ij}\}^{-1}.$$

As before, the corrected values $t + \delta t$ can be used as the basis of the new calculation and the process repeated to any desired accuracy. It is to be expected that the replacement of L_{ij} by its expected value Λ_{ij} will result in the approximation being less rapid; this is confirmed in particular examples by calculations given below, the difference being small, however.

The method has been used by Koshal (1933, 1939) in fitting a Pearson Type I curve,

$$y = y_0(x-\alpha)^{\mu_1}\,(\beta - x)^{\mu_2},$$

to a set of frequency data. First approximations to the maximum likelihood estimates of the four unknown parameters $\alpha, \beta, \mu_1, \mu_2$ (or $\theta_1, \theta_2, \theta_3, \theta_4$) were found by actually calculating a set of values of L and estimating its maximum position. The above method was then applied. In this case, if a typical group frequency is n and the expected proportion is p, we have, S denoting summation over the groups,

$$L = \text{constant} + Sn\log p, \quad L_i = S\frac{n}{p}p_i,$$

and

$$L_{ij} = -S\frac{n}{p^2}\,p_i p_j + S\frac{n}{p}p_{ij}.$$

As before, p_i denotes $\partial p/\partial\theta_i$ and p_{ij} denotes $\partial^2 p/\partial\theta_i\partial\theta_j$. The expected value of the last term is

$$NSp_{ij} = N\frac{\partial^2}{\partial\theta_i\partial\theta_j}Sp = 0,$$

where $N = Sn =$ total frequency, so that

$$\Lambda_{ij} = -NS\frac{p_i p_j}{p},$$

which were the values used by Koshal.

The covariance matrix* of the estimates T is approximately equal to $\{\Lambda_{ij}\}$ (Fisher, 1922); thus the variance of T_1 is approximately $-\lambda_{11}$ and the covariance of T_1 and T_2 is approximately $-\lambda_{12}$. The degree of approximation is such that terms of a higher order in $1/n$ are omitted.

To apply the method to dosage-mortality problems, suppose the probability of death is a function P of the dose (or log dose) x and of unknown parameters

* This has been used by Irwin & Cheeseman (1939) to derive the formulae for the variances and covariances of the estimates a and b of the parameters in the lethal dose distribution.

$\theta_1, \theta_2, \ldots$. The combined probability of the given set of survivors is

$$\Pi \frac{n!}{s!\,(n-s)!} P^{n-s} Q^s,$$

so that $$L = \text{const.} + S[(n-s)\log P + s\log Q],$$

and $$L_i = S\frac{n(Q-q)}{PQ} P_i,$$

where $q = s/n \doteq$ observed proportion of survivors; therefore

$$L_{ij} = S\left[\frac{n(Q-q)}{PQ} P_{ij} + nP_i P_j\left(-\frac{1}{P^2} + \frac{q}{P^2Q} - \frac{q}{PQ^2}\right)\right]$$

and $$\Lambda_{ij} = -S\frac{nP_iP_j}{PQ}.$$

If the distribution of lethal doses (or log lethal doses) depends only on parameters of scale and location, we have

$$P = P(\alpha + \beta x) = P(Y);$$

therefore $$P_\alpha = P'(\alpha+\beta x) = Z, \quad \text{and} \quad P_\beta = xZ,$$

where Z is the ordinate of the frequency distribution; thus

$$L_\alpha = S\frac{nZ}{PQ}(Q-q), \quad L_\beta = S\frac{xnZ}{PQ}(Q-q).$$

Putting $$\frac{nZ}{PQ}(Q-q) = \zeta = w\eta \quad \text{and} \quad \frac{nZ^2}{PQ} = w,$$

we have $$L_\alpha = S\zeta, \quad L_\beta = Sx\zeta.$$

Also $$L_{\alpha\alpha} = S\zeta', \quad L_{\alpha\beta} = Sx\zeta', \quad L_{\beta\beta} = Sx^2\zeta',$$

i.e. $$L_{\alpha\alpha} = S\frac{nZ(Q-q)}{PQ}\left(\frac{Z'}{Z} - \frac{Z}{P} + \frac{Z}{Q}\right) - S\frac{nZ^2}{PQ}$$

$$= -Sw + S\mu\zeta, \qquad \ldots\ldots(6)$$

where $$\mu = \frac{Z'}{Z} - \frac{Z}{P} + \frac{Z}{Q}.$$

Similarly $$L_{\alpha\beta} = -Swx + S\mu x\zeta, \qquad \ldots\ldots(7)$$

$$L_{\beta\beta} = -Swx^2 + S\mu x^2\zeta. \qquad \ldots\ldots(8)$$

The expected value of ζ is zero, so that

$$\Lambda_{\alpha\alpha} = -Sw, \quad \Lambda_{\alpha\beta} = -Swx, \quad \Lambda_{\beta\beta} = -Swx^2.$$

The equations for the corrections $\delta a, \delta b$ to approximations a_i, b_i to the maximum likelihood estimates, using the "expected" coefficients Λ, are thus

$$\delta a Sw + \delta b Swx = Sw\eta,$$
$$\delta a Swx + \delta b Swx^2 = Swx\eta,$$

i.e. the same as equations (2) and (3). Thus Fisher's exact maximum likelihood

method of correcting the regression line is exactly equivalent, in the case of any distribution defined by parameters of scale and position, to the above method of calculating successive approximations to the maximum likelihood solutions. In the case of the normal distribution, it should be noted that if a and b are the maximum likelihood estimates of α and β, then $\bar{x} = (5-a)/b$ and $s = 1/b$ are the maximum likelihood estimates of the mean and standard deviation m and σ, since it is easy to see that these values satisfy $\partial L/\partial m = 0$, $\partial L/\partial \sigma = 0$. In other words, the problem can equally be regarded as the estimation of the population mean and standard deviation from the data.

Furthermore, it may happen from some peculiarity of the experimental material that each experiment consists of one item, i.e. $n_1 = n_2 = \ldots = 1$, and the number of survivors is either 0 or 1. For example, the dose might represent some quantity which can be measured but not controlled.

Provided always there is independent evidence on which to base the assumption of the normal distribution (or of some other known form), there appears to be no reason why such data should not be effective for the purpose of drawing inferences about the population.

It is true that for the purpose of testing the hypothesis, say of normality, it will be necessary to group the data, and this will also be efficacious in obtaining a provisional probit line, i.e. first approximations to a and b; but for the exact estimation of the population parameters this is unnecessary (and would, in fact, result in a loss of information), for there is no difficulty in carrying out the calculations given above, or illustrated later in Tables II and III, with the values of q equal to 0 or 1. On the other hand, the exact problem cannot be regarded as one of fitting a line to the plotted probits, since all the latter are infinite in one direction or the other.

Another convenient way of regarding the problem of finding the maximum likelihood estimates is a geometrical one. We require to find the values a and b of α and β which are such that $L_\alpha = \partial L(\alpha,\beta)/\partial\alpha$ and $L_\beta = \partial L(\alpha,\beta)/\partial\beta$ are zero. Regarding α, β, γ as cartesian co-ordinates in three-dimensional space, we require to find the point $P(a, b, 0)$ where the two surfaces $\gamma = L_\alpha$, $\gamma = L_\beta$ and the plane $\gamma = 0$ meet. This is the same as the point of intersection of the curve $L_\alpha = 0$ in the plane $\gamma = 0$ (the horizontal plane) and the curve $L_\beta = 0$ in the same plane. If $P_1(a_1, b_1, 0)$ is an approximation to P, the tangent plane to the first surface at the point vertically above P_1 cuts the horizontal plane in a line which is near the first curve.

Using co-ordinates δa, δb with reference to P_1 as origin, this line has the equation

$$\frac{\partial L(a_1 b_1)}{\partial\alpha} + \delta a\,\frac{\partial^2 L(a_1 b_1)}{\partial\alpha^2} + \delta b\,\frac{\partial^2 L(a_1 b_1)}{\partial\alpha\,\partial\beta} = 0,$$

or in the simpler notation

$$L_\alpha + \delta a L_{\alpha\alpha} + \delta b L_{\alpha\beta} = 0.$$

Similarly there is a line near the second curve with the equation

$$L_\beta + \delta a L_{\alpha\beta} + \delta b L_{\beta\beta} = 0$$

and the intersection P_2 of these lines is a closer approximation to P than P_1. The values of $L_{\alpha\alpha}$, etc. are given in equations (6), (7) and (8) and the effect of replacing them by their expected values $\Lambda_{\alpha\alpha}$, etc. is to replace the tangent planes by planes through the points of contact differing slightly in direction, and to replace the above lines by lines whose equations are given by (2) and (3).

3. The goodness of fit test

In the regression line treatment of the problem, the test of the hypothesis of normality is provided by testing the residual variance χ^2 about the regression line, with degrees of freedom two less than the number of groups. From the point of view we are considering, it would appear more natural to calculate χ^2 as in the comparison of observed and expected frequencies, i.e.

$$\chi^2 = S\frac{(s-nP)^2}{nP} + S\frac{(n-s-nQ)^2}{nQ}$$

$$= S\frac{n(Q-q)^2}{PQ} = S\frac{nZ^2}{PQ}\left(\frac{Q-q}{Z}\right)^2 = Sw\eta^2,$$

with $k-2$ degrees of freedom.

The residual variance about the regression line is

$$Sw(\eta - \dot{x}\,\delta a - x\,\delta b)^2 = Sw\eta^2 - \delta a Sw\eta - \delta b Swx\eta,$$

so that the two values of χ^2 are identical when $\delta a = 0$, $\delta b = 0$, i.e. when the maximum likelihood estimates have been approached sufficiently closely.

4. Comparison of methods of solving maximum likelihood equations

For convenience we refer to the method using equations $\sum_{j=1}^{s} L_{ij}\delta t_j = -L_i$ as method I and that using the expected values of the coefficients, viz.

$$\sum_{j=1}^{s} \Lambda_{ij}\delta t_j = -L_i$$

as method II. Before comparing them arithmetically, it is of interest to enquire whether the two are ever identical, i.e. $L_{ij} = \Lambda_{ij}$, for a "scale-location" distribution. From (6), (7) and (8) this requires that

$$\mu \equiv \frac{Z'}{Z} - \frac{Z}{P} + \frac{Z}{Q} = 0.$$

Now $Z = dP/dY = P'$, so that the probability integral P must satisfy the differential equation

$$P'' - \frac{P'^2}{P} + \frac{P'^2}{1-P} = 0.$$

An integral of $$P'' + P'^2 f(P) = 0$$
is $$P' e^{\int f(P)\,dP} = C,$$
so that $$P' = CP(1-P),$$
and hence $$P = \frac{e^{B+CY}}{1+e^{B+CY}}.$$

Since $Y = \alpha + \beta x$, we can choose values $B = 0$, $C = 2$ for the arbitrary constants, giving

$$P = \frac{e^{2Y}}{1+e^{2Y}},$$

and ordinate $Z = P' = \frac{1}{2}\operatorname{sech}^2 Y$ which is the distribution given by Fisher & Yates (1938). Thus for this distribution $\mu = 0$ and $L_{\alpha\alpha} = \Lambda_{\alpha\alpha}$, $L_{\alpha\beta} = \Lambda_{\alpha\beta}$, $L_{\beta\beta} = \Lambda_{\beta\beta}$; hence equations (2) and (3) will give the most rapid solution.

Tables for facilitating these calculations have been given by Fisher & Yates (1938); it would appear that Tables XII–XIV of this work supply similar tables for fitting a distribution of the type $P = \sin^2\phi = \sin^2(\alpha+\beta x)$; here

$$\mu = -2\cot 2\phi, \quad w = 4n, \quad \zeta = 4n(Q-q)\operatorname{cosec} 2\phi,$$

so that $$L_{\alpha\alpha} = -4N - 8Sn(Q-q)\cot 2\phi \operatorname{cosec} 2\phi, \text{ etc.},$$

and any advantage in rapidity gained by using method I is almost certainly offset by the simplicity in method II, since $\Lambda_{\alpha\alpha} = -4N$, $\Lambda_{\alpha\beta} = -4NSx$, $\Lambda_{\beta\beta} = -4NSx^2$. The point has not been tested, since no examples have come to hand in which a $P = \sin^2\phi$ distribution has been envisaged.

To apply method I to the normal distribution we have ordinate

$$Z = \frac{1}{\sqrt{2}\,\pi} e^{-\frac{1}{2}(Y-5)^2};$$

therefore $$\mu = 5 - Y + \frac{Z}{Q} - \frac{Z}{P}.$$

The two methods of successive approximation have been applied to each of the following three sets of data taken more or less at random from those already published for illustrative purposes.

Example (i). Antipneumococcus serum given to five groups of forty mice; Wilson Smith's data (Irwin & Cheeseman, 1939, p. 179).

Serum dose c.c.	x	Deaths out of 40
0·000625	-2	33
0·00125	-1	22
0·0025	0	8
0·005	1	5
0·01	2	2

Since the increased dose resulted in more survivors, we call q the proportion of deaths.

Example (ii). Mice injected with *Bact. Typhi-murium*, sample A; Topley's data (Irwin & Cheeseman, 1939, p. 180).

Dose (mg.)	x	Survivors out of 5
0·0625	−3	4
0·125	−2	3
0·25	−1	2
0·5	0	0
1·0	1	0
2·0	2	0
4·0	3	0

Example (iii). Brine shrimps, *Artemia salina*, in arsenical solutions having concentrations in geometrical progression (Fisher & Yates, 1938, p. 5).

Solution	x	Survivors out of 8
C	−3	8
D	−2	8
E	−1	6
F	0	5
G	1	5
H	2	1
I	3	0

In each case the scale of x has been chosen with a central origin for convenience. The last example was used by Fisher & Yates (1938) to illustrate the use of a table (Table XI) drawn up for solving these problems; as, however, the arithmetical work in this note has, for purposes of comparison, been taken to more places of decimals than practical work demands, this table has not been used, and the requisite areas and ordinates of the normal curve have been taken from tables of the normal curve (Pearson, 1930). The results of the comparison are shown in Table I.

Most of the values given are probably exact; there may, however, be an occasional error of one unit in the last place. The first approximation in Example (i) is one used by Irwin & Cheeseman; as it has already been calculated from the data, the errors in the coefficients are smaller than in Examples (ii) and (iii). In Example (ii) the first approximation was found by fitting a line by eye to the observed probits, and in Example (iii) the first approximation was found by a

Table I. *Comparison of methods of successive approximation to solutions of maximum likelihood equations*

Example	Approximation		Successive corrections to a and b			
	First	Final	Method II		Method I	
			δa	δb	δa	δb
(i)	$Y=5{\cdot}5451+0{\cdot}6689x$	$Y=5{\cdot}5544+0{\cdot}6761x$	0·0095 −0·0002	0·0074 −0·0002	0·0093 0·0000	0·0071 0·0001
(ii)	$Y=7+x$	$Y=6{\cdot}5168+0{\cdot}8635x$	−0·6475 0·1350 0·0267 0·0024 0·0002	−0·2034 0·0530 0·0127 0·0011 0·0001	−0·6133 0·1190 0·0110 0·0001 —	−0·1952 0·0533 0·0053 0·0001 —
(iii)	$Y=4{\cdot}6+0{\cdot}6x$	$Y=4{\cdot}5658+0{\cdot}7128x$	−0·0259 −0·0080 −0·0003 —	0·0891 0·0215 0·0021 0·0001	−0·0268 −0·0072 0·0002 —	0·0984 0·0142 0·0002 —

Method II. Expected values of $\partial^2 L/\partial \alpha^2$, etc. used.
Method I. Actual values of $\partial^2 L/\partial \alpha^2$, etc. used.

preliminary regression calculation with the coefficients rounded off to one place of decimals. The arithmetical details of the calculations (Example (ii)) of the 2nd approximation by the two methods are given for illustration in Tables II and III.

It is seen from a comparison of the results that there is a slight advantage, from the point of view of rapidity, in method I. On the other hand, method II entails a little less arithmetical work, and if Fisher & Yates' tables are used it is probable that this advantage would be greater, although the point has not been investigated.

Table II. *Typical calculation of corrections to estimates of parameters, Example (ii), method I*

1st approximation, $Y = 7 + x$

Dose (mg.)	x	n	s	q	$t = Y-5$	P*	Q*	Z*	$Q-q$	$\frac{Z}{PQ}$	$\zeta = \frac{Z}{PQ}(Q-q)$†	$\frac{Z(P-Q)}{PQ} - t = \mu$	$\zeta' = \mu\zeta - \frac{Z^2}{PQ}$
0·0625	−3	5	4	0·8	−1	0·15866	0·84134	0·24197	0·04134	1·8126	0·07493	−0·2374	−0·45638
0·125	−2	5	3	0·6	0	0·50000	0·50000	0·39894	−0·10000	1·5958	−0·15958	0·0000	−0·63663
0·25	−1	5	2	0·4	1	0·84134	0·15866	0·24197	−0·24134	1·8126	−0·43745	0·2374	−0·54245
0·5	0	5	0	0·0	2	0·97725	0·02275	0·05399	0·02275	2·4285	0·05525	0·3180	−0·11355
1·0	1	5	0	0·0	3	0·99865	0·00135	0·00443	0·00135	3·2874	0·00444	0·2785	−0·01333
2·0	2	5	0	0·0	4	0·99997	0·00003	0·00013	0·00003	4·2200	0·00013	0·2200	−0·00054
4·0	3	5	0	0·0	5	1·00000	0·00000	0·00000	0·00000	5·1900	0·00000	0·1900	−0·00001

$$L_\alpha = S\zeta = -0{\cdot}46228, \quad L_\beta = Sx\zeta = 0{\cdot}53652$$

$$L_{\alpha\alpha} = S\zeta' = -1{\cdot}7629, \quad L_{\alpha\beta} = Sx\zeta' = 3{\cdot}1704, \quad L_{\beta\beta} = Sx^2\zeta' = -7{\cdot}2120$$

$$\delta a = \frac{L_\beta L_{\alpha\beta} - L_\alpha L_{\beta\beta}}{L_{\alpha\alpha} L_{\beta\beta} - L_{\alpha\beta}^2} = -0{\cdot}6133, \quad \delta b = \frac{L_\alpha L_{\alpha\beta} - L_\beta L_{\alpha\alpha}}{L_{\alpha\alpha} L_{\beta\beta} - L_{\alpha\beta}^2} = -0{\cdot}1952$$

$$a = 7 \qquad\qquad b = 1$$

$$a + \delta a = 6{\cdot}3867 \qquad\qquad b + \delta b = 0{\cdot}8048$$

2nd approximation, $Y = 6{\cdot}3867 + 0{\cdot}8048x$. Final approximation, $Y = 6{\cdot}5168 + 0{\cdot}8635x$.

Table III. *Typical calculation of corrections to estimates of parameters, Example (ii), method II*

1st approximation, $Y = 7 + x$

Dose (mg.)	x	n	s	q	$t = Y-5$	P*	Q*	Z*	$Q-q$	$\eta = \frac{Q-q}{Z}$	w†	wx†
0·0625	−3	5	4	0·8	−1	0·15866	0·84134	0·24197	0·04134	0·1708	0·4386	−1·3158
0·125	−2	5	3	0·6	0	0·50000	0·50000	0·39894	−0·10000	−0·2507	0·6368	−1·2732
0·25	−1	5	2	0·4	1	0·84134	0·15866	0·24197	−0·24134	−0·9974	0·4386	−0·4386
0·5	0	5	0	0·0	2	0·97725	0·02275	0·05399	0·02275	0·4214	0·1311	0·0000
1·0	1	5	0	0·0	3	0·99865	0·00135	0·00443	0·00135	0·3046	0·0146	0·0146
2·0	2	5	0	0·0	4	0·99997	0·00003	0·00013	0·00003	0·2369	0·0006	0·0012
4·0	3	5	0	0·0	5	1·00000	0·00000	0·00000	0·00000	0·1928	0·0000	0·0000

$$Swx\eta = 0{\cdot}5366 \qquad Swx^2 = 6{\cdot}9494 \qquad Sw = 1{\cdot}6601 \qquad Swx = -3{\cdot}0118$$

$$-\bar{x}Sw\eta = -0{\cdot}8387 \qquad -\bar{x}Swx = 5{\cdot}4640 \qquad \bar{x} = -1{\cdot}8142$$

$$Sw\eta(x-\bar{x}) = -0{\cdot}3021 \qquad Sw(x-\bar{x})^2 = 1{\cdot}4854 \qquad Sw\eta = -0{\cdot}4623$$

$$\delta b = \frac{Sw\eta(x-\bar{x})}{Sw(x-\bar{x})^2} = -0{\cdot}2034 \qquad \bar{\eta} = -0{\cdot}2785$$

$$-\bar{x}\delta b = -0{\cdot}3690$$

$$b = 1 \qquad \delta a = -0{\cdot}6475$$

$$b + \delta b = 0{\cdot}7966 \qquad a = 7$$

$$a + \delta a = 6{\cdot}3525$$

2nd approximation, $Y = 6{\cdot}3525 + 0{\cdot}7966x$. Final approximation, $Y = 6{\cdot}5168 + 0{\cdot}8635x$.

* Five significant figures were used for P, Q and Z, where possible, but only five places of decimals are shown in the table; similarly in Table III.

† As $n = 5$ in each sample, it has been omitted for convenience from ζ, ζ' and from w, wx in Table III.

5. Summary

The usual practical maximum likelihood treatment of dosage-mortality problems (consisting of the transformation of percentage surviving into probits, adjustment and weighting of the latter, and calculation of successive regression lines) is shown to be equivalent to calculating successive corrections to the regression coefficients. The process is exactly equivalent to the method, given elsewhere by Fisher, of obtaining the maximum likelihood estimates of the parameters defining the distribution. A refinement of this method, using the actual values of the second derivatives of the likelihood function, instead of the expected values, converges a little more rapidly when applied to the normal distribution, but this advantage is offset by some extra arithmetical work. The two methods are exactly equivalent only for the distribution specified by $P = \frac{1}{2}\operatorname{sech}^2 z$.

The writer is greatly indebted to Mr E. D. van Rest and to Prof. R. A. Fisher for much useful help and advice.

REFERENCES

Bliss, C. I. (1935). *Ann. Appl. Biol.* **22**, 134.

—— (1938). *Quart. J. Pharm.* **11**, 192.

Fisher, R. A. (1922). *Philos. Trans.* A, **222**, 309.

Fisher, R. A. & Yates, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh & London: Oliver & Boyd.

Irwin, J. O. & Cheeseman, E. A. (1939). *J. R. Statist. Soc.* Suppl. **6**, 174.

Koshal, R. S. (1933). *J. R. Statist. Soc.* **96**, 303.

—— (1939). *Ann. Eugen., Lond.*, **9**, 209.

Pearson, K. (1930). *Tables for Statisticians and Biometricians*, Part I, 3rd ed., Biometrika Office.

A NOTE ON FURTHER PROPERTIES OF STATISTICAL TESTS

By E. S. PEARSON

Dr P. L. Hsu has suggested that I should write a short introductory note on the origin of the idea involved in his paper and in that of Dr Simaika's which follows.* In searching some twelve years ago for a systematic method of choosing the best test of a statistical hypothesis H_0, Prof. Neyman and I came to the conclusion that an essential preliminary to any mathematical formulation of the problem was the definition of a set of admissible alternative hypotheses, $C(H)$. Starting from this viewpoint, our first method of selecting a test involved the use of the likelihood ratio, but, however useful as a practical method of attack, the principle underlying this approach was somewhat arbitrary. A more fundamental procedure, later developed, was to choose a test paying regard to its power function, that is to say, to the chance that its use would lead to the rejection of H_0 if an alternative $H \neq H_0$ of $C(H)$ were true. It then appeared that a number of statistical tests in common use had the remarkable property that they maximized this chance for every alternative to H_0 in $C(H)$. Such tests were termed uniformly most powerful tests of H_0 with regard to $C(H)$.

That there were limitations to the situations in which a uniformly most powerful test could exist soon, however, became clear. These limitations were gradually explored, and the following papers are further contributions to the subject. It was found that these tests generally, though not always, concerned the value of a single parameter. Such are tests of the hypothesis that a mean or a standard deviation has a specified value, or that the difference between two means or two standard deviations is zero. Further, in these cases the class of alternatives must be restricted; thus the two-sample t-test of the hypothesis that two population means ξ_1 and ξ_2 are equal, is only uniformly most powerful for the situation in which the alternatives considered are defined by $\xi_1-\xi_2 \geqslant 0$ or by $\xi_1-\xi_2 \leqslant 0$ but not for both at the same time.

In this connexion, in 1935, Kołodziejczyk was able to prove that for tests of a linear hypothesis, no uniformly most powerful test could exist if the number of parameters involved was greater than unity. This result was important, since the majority of tests used in the analysis of variance can be reduced to tests of a linear hypothesis.

This limitation of tests regarding the value of two or more parameters can be illustrated by a geometric presentation. Since the most important features of the problem can be illustrated when H_0 is a simple hypothesis concerning the value of two parameters, I shall take this case, using notation already adopted in this connexion.

* See pp. 62–69 and pp. 70–80 below.

Suppose that the elementary probability law of random variables $x_1, \ldots, x_n$, whose particular values are given by observation, is of form

$$p(x_1, \ldots, x_n \mid \theta_1, \theta_2) = p(E \mid \theta_1, \theta_2), \tag{1}$$

θ_1, θ_2 being the two population parameters. For a critical region w of size α associated with a given test, we may write

$$P\{E\epsilon w \mid \theta_1, \theta_2\} = \int \ldots \int_w p(E \mid \theta_1, \theta_2)\, dx_1 \ldots dx_n = \beta(\theta_1, \theta_2 \mid w). \tag{2}$$

If the hypothesis H_0 which w has been selected to test assumes that

$$\theta_1 = \theta_1^0, \quad \theta_2 = \theta_2^0, \tag{3}$$

then
$$\beta(\theta_1^0, \theta_2^0 \mid w) = \alpha, \tag{4}$$

where α is the significance level chosen.

A power surface may be obtained by taking rectangular axes for θ_1 and θ_2 in a horizontal plane and plotting $\beta(\theta_1, \theta_2 \mid w)$ as a vertical ordinate. If w_0 were a critical region associated with a uniformly most powerful test of H_0, then its power surface would fall nowhere below the surfaces derived from other critical regions satisfying (4). No unique surface with this property will, however, in general exist. If, for instance, we choose w_0 so that the surface will rise quickly in the direction parallel to the axis of θ_1, we shall reduce the rate of increase in the direction of θ_2, and vice versa. Power surfaces of alternative critical regions may, in fact, cross one another in a complicated way, but no single surface can everywhere lie above all others. If we confine attention to tests for which the power surface has a minimum ordinate of α at the point θ_1^0, θ_2^0, i.e. to unbiased tests of H_0, we shall still be unable to find a uniformly most powerful test in this restricted field.

The difficulty in choice between alternative tests can, indeed, only be solved by a further formulation of the requirements of a satisfactory test. Several lines of attack are open:

(i) To lay down conditions for the form of the power surface in the neighbourhood of the point θ_1^0, θ_2^0. Here we may describe the objective as to make as large as possible the chance of detecting small departures in θ_1 and θ_2 from the values specified by H_0. A method of approaching the problem from this point of view leads to the development of the unbiased test of Type C (Neyman & Pearson, 1938).

(ii) To regard it as of more importance to control the form of the power surface at some distance from its minimum point; for example, to try to select a critical region for which the power surface reaches the level

$$\beta(\theta_1, \theta_2 \mid w) = 0{\cdot}95, \tag{5}$$

along a contour lying inside the corresponding contour associated with any other test. This method of approach has been examined by Dr B. L. Welch,

but his results are not yet published. It is possible that methods (i) and (ii) will lead to the same result.

(iii) To consider whether from the practical point of view, if H_0 is not true, the importance of the departure of the unknown parameters from θ_1^0, θ_2^0 can be measured by a single parameter,

$$\lambda = f(\theta_1, \theta_2). \tag{6}$$

If this is so, we are in fact defining a system of contours on the θ_1, θ_2 plane along any one of which we should like the ordinates of the power surface to be constant. Such a system would be defined, for instance, by

$$\lambda^2 = (\theta_1 - \theta_1^0)^2 + (\theta_2 - \theta_2^0)^2, \tag{7}$$

and if $\beta(\theta_1, \theta_2 \mid w)$ is to be constant for values of θ_1, θ_2 satisfying (7), the contours of the power surface will be circles of radius λ. Alternative tests would then be confined to those whose power surfaces had circular contours, H_0 would be the hypothesis that $\lambda = 0$ and the uniformly most powerful test, if it exists, would be that for which

$$\beta(\lambda \mid w_0) \geqslant \beta(\lambda \mid w) \tag{8}$$

for $\lambda > 0$ and all alternative critical regions w satisfying the conditions stipulated.

The problem thus presented in the case of a simple hypothesis concerning two parameters will arise in similar form when H_0 is composite and concerns the value of many parameters $\theta_1, \theta_2, \ldots, \theta_c$. In a number of multivariate problems we have reached a position in which:

(*a*) tests of statistical hypotheses concerning the values of several population parameters have been derived, as well as their power functions;

(*b*) these power functions have been shown to depend on the value of a single function

$$\lambda = f(\theta_1, \theta_2, \ldots, \theta_c)$$

of the parameters considered.

In the following contributions Dr Hsu and Dr Simaika have examined three of these tests, that concerned with the general linear hypothesis, with Hotelling's generalized T^2 and with the multiple correlation coefficient. They have shown that of tests whose power function depends only on a certain function λ of the population parameters, the existing tests are the uniformly most powerful. It is of course true that in the problems in question no alternative tests are at present available or indeed likely to become so. Nevertheless, I believe that the discovery, resulting from Dr Hsu's initiative, of the relationship between the test function and a corresponding comprehensive collective character in the population, has taken us a step farther in our understanding of the properties of statistical tests. Further, this relationship between E^2 and λ, T^2 and ψ^2, D^2 and Δ^2, R^2 and ρ^2 seems to lead us round by another route to the problem of statistical estimation.

REFERENCES

KOŁODZIEJCZYK, ST (1935). *Biometrika*, **27**, 161.
NEYMAN, J. & PEARSON, E. S. (1938). *Statist. Res. Mem.* **2**, 25.

ANALYSIS OF VARIANCE FROM THE POWER FUNCTION STANDPOINT

By P. L. HSU

A FRESH study on the classical analysis of variance tests in the light of the Neyman-Pearson theory was started by Kołodziejczyk (1935), who formulated the class of linear hypotheses for which these tests may be employed. As a linear hypothesis is defined relative to the set of admissible hypotheses, the study of the E^2-test (by which we denote any test falling under the usual methods of analysis of variance) may be made with reference to its power function. P. C. Tang (1938) showed how the power function was related to R. A. Fisher's (C) distribution (Fisher, 1928) and so was able to appraise the chance of detecting the falsehood of a linear hypothesis using the E^2-test. The great theoretical value of the power function lies, however, in its use in comparing the relative merits of alternative tests of the same hypothesis. In this paper we shall prove a theorem (p. 63) which asserts that out of a certain class of tests the E^2-test is uniformly most powerful.

In his paper Tang has used an orthogonal transformation in the sample space which enabled the general linear hypothesis to be reduced to the following simple form: Given the elementary probability law

$$p(y_1, \ldots, y_m, z_1, \ldots, z_n) = \{\sqrt{(2\pi)}\,\sigma\}^{-(m+n)} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m} (y_i-\eta_i)^2 + \sum_{i=1}^{n} z_i^2\right)\right\}, \tag{1}$$

where all real values of $\eta_1, \ldots, \eta_m$ and all positive values of σ are admissible, the hypothesis is that n_1 ($\leqslant m$) of the η's have the true value 0:

$$\eta_1 = \eta_2 = \ldots = \eta_{n_1} = 0. \tag{2}$$

We call the above hypothesis H_0.

We shall set

$$E^2 = \sum_{i=1}^{n_1} y_i^2 \Big/ \left(\sum_{i=1}^{n_1} y_i^2 + \sum_{i=1}^{n} z_i^2\right), \tag{3}$$

and call w_0 (of size ϵ) the critical region for the rejection of H_0 defined by the inequality

$$E^2 \geqslant E_\epsilon^2, \tag{4}$$

where E_ϵ^2 is a constant so determined that the probability that (4) is true, given that (2) is true, equals ϵ.

The power function of w_0 as given by Tang can be written

$$e^{-\lambda} \sum_{h=0}^{\infty} \{h!\,B(\tfrac{1}{2}n_1+h, \tfrac{1}{2}n)\}^{-1} \lambda^h \int_{E_\epsilon^2}^{1} (E^2)^{\frac{1}{2}n_1+h-1} (1-E^2)^{\frac{1}{2}n-1}\, d(E^2), \tag{5}$$

where
$$\lambda = \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} \eta_i^2. \tag{6}$$

An outstanding feature of the power function (5) is that it depends on the single parameter λ. Our problem is, does there exist another critical region of size ϵ whose power function depends on λ alone and which is more powerful than w_0 for certain values of λ? The answer is contained in the following theorem and is in the negative.

THEOREM. *Suppose that the critical region w satisfies the following conditions:*

(*a*) *w is of size ϵ,*

(*b*) *the power function of w depends on the single parameter λ.*

Let $\beta(\lambda)$ be the power function of w and $\beta_0(\lambda)$ be the power function (5) *of w_0. Then*
$$\beta(\lambda) \leqslant \beta_0(\lambda) \tag{7}$$
for all positive values of λ.

Proof. In the place of $z_1, \ldots, z_n$ we substitute spherical co-ordinates, viz. the radius vector $r = (\Sigma z^2)^{\frac{1}{2}}$ and $n-1$ angles, $\theta_1, \ldots, \theta_{n-1}$. We deduce from (1) that
$$p(y_1, \ldots, y_m, \theta_1, \ldots, \theta_{n-1}, r) = p(y_1, \ldots, y_{n_1})\, p(y_{n_1+1}, \ldots, y_m)\, p(\theta_1, \ldots, \theta_{n-1})\, p(r), \tag{8}$$
where
$$p(y_1, \ldots, y_{n_1}) = \{\sqrt{(2\pi)}\,\sigma\}^{-n_1} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n_1} (y_i - \eta_i)^2\right\}, \tag{9}$$
$$p(y_{n_1+1}, \ldots, y_m) = \{\sqrt{(2\pi)}\,\sigma\}^{-(m-n_1)} \exp\left\{-\tfrac{1}{2} \sum_{i=n_1+1}^{m} (y_i - \eta_i)^2\right\}, \tag{10}$$
$$p(r) = 2^{-\frac{1}{2}(n-2)}\, \sigma^{-n} (\Gamma \tfrac{1}{2} n)^{-1}\, r^{n-1} \exp\left(-\frac{1}{2\sigma^2} r^2\right), \tag{11}$$
and $p(\theta_1, \ldots, \theta_{n-1})$ is the well-known product of cosines which involves none of the parameters $\eta_1, \ldots, \eta_m$ and σ.

We now make the following successive transformations:
$$r = s^{\frac{1}{2}}, \quad s = t - \sum_{i=1}^{n_1} y_i^2, \quad y_i = t^{\frac{1}{2}} u_i \quad (i = 1, \ldots, n_1), \tag{12}$$
and also write
$$\gamma_i = \sigma^{-1} \eta_i \quad (i = 1, \ldots, n_1). \tag{13}$$
It follows that
$$p(y_{n_1+1}, \ldots, y_m, \theta_1, \ldots, \theta_{n-1}, u_1, \ldots, u_{n_1}, t) = p(y_{n_1+1}, \ldots, y_m)\, p(\theta_1, \ldots, \theta_{n-1})\, p(u_1, \ldots, u_{n_1}, t), \tag{14}$$
where
$$p(u_1, \ldots, u_{n_1}, t) = (\sqrt{2}\,\sigma)^{-(n+n_1)} \pi^{-\frac{1}{2}n_1} (\Gamma \tfrac{1}{2} n)^{-1} e^{-\lambda} t^{\frac{1}{2}(n+n_1-2)} \exp\left(-\frac{t}{2\sigma^2}\right) \times \left(1 - \sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\frac{\sqrt{t}}{\sigma} \sum_{i=1}^{n_1} \gamma_i u_i\right). \tag{15}$$

From now on we shall write y for the set of variables $y_{n_1+1}, \ldots, y_m$ and dy for $dy_{n_1+1}, \ldots, dy_m$, and use similar abbreviations θ, u, $d\theta$ and du.

Suppose now that the critical region w satisfies the conditions (a) and (b). Let $\Gamma(y, \theta, u, t)$ be the characteristic function of w, i.e. $\Gamma(y, \theta, u, t) = 1$ or 0 according as the sample point falls within w or not. Let W be the sample space. Then the power function of w is

$$\beta(\lambda) = \int_W \Gamma(y, \theta, u, t)\, p(y, \theta, u, t)\, dy\, d\theta\, du\, dt, \tag{16}$$

whence,

$$(\sqrt{2}\,\sigma)^{-(n+n_1)} \int_W \Gamma(y, \theta, u, t)\, p(y)\, p(\theta)\, t^{\frac{1}{2}(n+n_1-2)} \times \exp\left(-\frac{t}{2\sigma^2}\right)\left(1 - \sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} dy\, d\theta\, du\, dt = \pi^{\frac{1}{2}n_1}\Gamma(\tfrac{1}{2}n)\, \epsilon, \tag{17}$$

$$(\sqrt{2}\,\sigma)^{-(n+n_1)} \int_W \Gamma(y, \theta, u, t)\, p(y)\, p(\theta)\, t^{\frac{1}{2}(n+n_1-2)} \times \exp\left(-\frac{t}{2\sigma^2}\right)\left(1 - \sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \cdot \exp\left(\frac{\sqrt{t}}{\sigma} \sum_{i=1}^{n_1} \gamma_i u_i\right) dy\, d\theta\, du\, dt$$
$$= \pi^{\frac{1}{2}n_1}\Gamma(\tfrac{1}{2}n)\, e^{\lambda}\beta(\lambda) = F(\lambda) = F\left(\sum_{i=1}^{n_1} \gamma_i^2\right), \text{ say.} \tag{18}$$

Let W_1 be the sample space of θ, u and t, and put

$$(\sqrt{2}\,\sigma)^{-(n+n_1)} \int_{W_1} \Gamma(y, \theta, u, t)\, p(\theta)\, t^{\frac{1}{2}(n+n_1-2)} \times \exp\left(-\frac{t}{2\sigma^2}\right)\left(1 - \sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\frac{\sqrt{t}}{\sigma} \sum_{i=1}^{n_1} \gamma_i u_i\right) d\theta\, du\, dt - F(\lambda) = \phi(y, \gamma, \sigma). \tag{19}$$

Then, by (18),
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y)\, p(y)\, dy = 0, \tag{20}$$

i.e.
$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y_{n_1+1}, \ldots, y_m, \gamma_1, \ldots, \gamma_{n_1}, \sigma) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=n_1+1}^{m} y_i^2\right) \times \exp\left(\sum_{i=n_1+1}^{m} \alpha_i y_i\right) dy_{n_1+1} \ldots dy_m = 0, \tag{21}$$

on writing α_i for $(2\sigma^2)^{-1}\eta_i$ $(i = n_1 + 1, \ldots, m)$.

Equation (20) must hold true for all real values of the α's. Hence it follows from the well-known theorem on Laplace transformation* that

$$\phi(y, \gamma, \sigma) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=n_1+1}^{m} y_i^2\right) = 0,$$

* Cf. Doetsch (1937), p. 35, Theorem 1. Though the theorem referred to is stated for the case where the number of y's is one, it may easily be extended to the case of more than one y by induction.

i.e. $\phi(y,\gamma,\sigma) = 0$, whence, by (19),

$$(\sqrt{2}\,\sigma)^{-(n+n_1)}\int_{W_1}\Gamma(y,\theta,u,t)\,p(\theta)\,t^{\frac{1}{2}(n+n_1-2)} \times\exp\left(-\frac{t}{2\sigma^2}\right)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}\exp\left(\frac{\sqrt{t}}{\sigma}\sum_{i=1}^{n_1}\gamma_i u_i\right)d\theta\,du\,dt = F\left(\sum_{i=1}^{n_1}\gamma_i^2\right). \quad (22)$$

In particular, from $\phi(y,0,\sigma) = 0$ and (17) we have

$$(\sqrt{2}\,\sigma)^{-(n+n_1)}\int_{W_1}\Gamma(y,\theta,u,t)\,p(\theta)\,t^{\frac{1}{2}(n+n_1-2)} \times\exp\left(-\frac{t}{2\sigma^2}\right)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}d\theta\,du\,dt = \pi^{\frac{1}{2}n_1}\Gamma(\tfrac{1}{2}n)\,\epsilon. \quad (23)$$

Letting W_2 be the sample space of θ and u, we get respectively from (23) and (22) that

$$(\sqrt{2}\,\sigma)^{-(n+n_1)}\int_0^\infty t^{\frac{1}{2}(n+n_1-2)} \times\exp\left(-\frac{t}{2\sigma^2}\right)dt\int_{W_2}\Gamma(y,\theta,u,t)\,p(\theta)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}d\theta\,du = \pi^{\frac{1}{2}n_1}\Gamma(\tfrac{1}{2}n)\,\epsilon, \quad (24)$$

$$(\sqrt{2}\,\sigma)^{-(n+n_1)}\int_0^\infty t^{\frac{1}{2}(n+n_1-2)}\exp\left(-\frac{t}{2\sigma^2}\right)dt\int_{W_2}\Gamma(y,\theta,u,t)\,p(\theta)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)} \times\exp\left(\frac{\sqrt{t}}{\sigma}\sum_{i=1}^{n_1}\gamma_i u_i\right)d\theta\,du = F\left(\sum_{i=1}^{n_1}\gamma_i^2\right). \quad (25)$$

Hence, on developing the left-hand side of (25) into power series in the γ's, we must have

$$\int_0^\infty t^{\frac{1}{2}(n+n_1-2+h)}\exp\left(-\frac{t}{2\sigma^2}\right)dt \times\int_{W_2}\Gamma(y,\theta,u,t)\,p(\theta)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}\left(\sum_{i=1}^{n_1}\gamma_i u_i\right)^h d\theta\,du = 0 \text{ for odd } h, \quad (26)$$

$$2^{-\frac{1}{2}(n+n_1)}\sigma^{-(n+n_1+2h)}\int_0^\infty t^{\frac{1}{2}(n+n_1-2)+h} \times\exp\left(-\frac{t}{2\sigma^2}\right)dt\int_{W_2}\Gamma(y,\theta,u,t)\,p(\theta)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}\left(\sum_{i=1}^{n_1}\gamma_i u_i\right)^{2h}d\theta\,du = a_h\left(\sum_{i=1}^{n}\gamma_i^2\right)^h \quad (h=1,2,3,\ldots). \quad (27)$$

Further, equations (24) and (27) may be written as

$$\int_0^\infty t^{\frac{1}{2}(n+n_1-2)}\exp\left(-\frac{t}{2\sigma^2}\right) \times\left[\int_{W_2}\Gamma(y,\theta,u,t)\,p(\theta)\left(1-\sum_{i=1}^{n_1}u_i^2\right)^{\frac{1}{2}(n-2)}d\theta\,du-\frac{\pi^{\frac{1}{2}n_1}\Gamma(\frac{1}{2}n)}{\Gamma\frac{1}{2}(n+n_1)}\epsilon\right]dt = 0, \quad (28)$$

$$\int_0^\infty t^{\frac{1}{2}(n+n_1-2)} \exp\left(-\frac{t}{2\sigma^2}\right)\left[\int_{W_1}' \Gamma(y,\theta,u,t)\, p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)}\right.$$

$$\left.\times\left(\sum_{i=1}^{n_1}\gamma_i u_i\right)^{2h} d\theta\, du - \frac{a_h\left(\sum_{i=1}^{n_1}\gamma_i^2\right)^h}{2^h\,\Gamma\{\frac{1}{2}(n+n_1)+h\}}\right] dt = 0 \quad (h=1,2,3,\ldots). \quad (29)$$

Equations (28), (26) and (29) must hold true for all positive values of σ. Hence, by the theorem of Laplace transformation,* the functions within the square brackets in (28) and (29) and the inner integral in (26) must vanish identically:

$$\int_{W_1} \Gamma(y,\theta,u,t)\, p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} d\theta\, du = \frac{\pi^{\frac{1}{2}n_1}\Gamma(\frac{1}{2}n)}{\Gamma\frac{1}{2}(n+n_1)}\epsilon, \quad (30)$$

$$\int_{W_1} \Gamma(y,\theta,u,t)\, p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)}\left(\sum_{i=1}^{n_1}\gamma_i u_i\right)^h d\theta\, du = 0 \text{ for odd } h, \quad (31)$$

$$\int_{W_1} \Gamma(y,\theta,u,t)\, p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)}\left(\sum_{i=1}^{n_1}\gamma_i u_i\right)^{2h} d\theta\, du$$

$$= \frac{a_h\left(\sum_{i=1}^{n_1}\gamma_i^2\right)^h}{2^h\,\Gamma\{\frac{1}{2}(n+n_1)+h\}} \quad (h=1,2,3,\ldots). \quad (32)$$

From (31) and (32) we infer that

$$\int_{W_1} \Gamma(y,\theta,u,t)\, p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\sum_{i=1}^{n_1}\gamma_i u_i\right) d\theta\, du = G\left(\sum_{i=1}^{n_1}\gamma_i^2\right). \quad (33)$$

Now for any given values of y and t the integral $\int_{W_1}\Gamma(y,\theta,u,t)\, f(\theta,u)\, d\theta\, du$ equals $\int_{w_2} f(\theta,u)\, d\theta du$, where w_2 is the set of points in the sample space of θ and u for which $\Gamma(y,\theta,u,t)=1$. Hence (30) and (33) are equivalent to

$$\int_{w_2} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} d\theta\, du = \frac{\pi^{\frac{1}{2}n_1}\Gamma(\frac{1}{2}n)}{\Gamma\frac{1}{2}(n+n_1)}\epsilon, \quad (34)$$

$$\int_{w_2} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\sum_{i=1}^{n_1}\gamma_i u_i\right) d\theta\, du = G\left(\sum_{i=1}^{n_1}\gamma_i^2\right). \quad (35)$$

The conditions (34) and (35) are necessary and sufficient that the critical region w should have the properties (*a*) and (*b*).

On the other hand, from (12) we have

$$E^2 = \sum_{i=1}^{n_1} u_i^2; \quad (36)$$

hence w_0 is the region defined by the inequality

$$\sum_{i=1}^{n_1} u_i^2 \geqslant E_\epsilon^2. \quad (37)$$

* Cf. footnote, p. 64.

Since w_0 is of size ϵ, we must get the same right-hand side of (34) when in the left-hand side we substitute w_0 for w_2. Hence

$$\int_{w_2} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} d\theta\, du = \int_{w_0} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} d\theta\, du. \tag{38}$$

Let

$$\int_{w_0} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\sum_{i=1}^{n_1} \gamma_i u_i\right) d\theta\, du = G_0\left(\sum_{i=1}^{n_1} \gamma_i^2\right). \tag{39}$$

With the help of (37) and (38) we may now appeal to the lemma proved in the Appendix and conclude that

$$G(\lambda) \leqslant G_0(\lambda), \tag{40}$$

whence, replacing γ_i by $\sigma^{-1}\sqrt{t}\gamma_i$ in the integrals in (35) and (39),

$$\int_{w_2} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\frac{\sqrt{t}}{\sigma}\sum_{i=1}^{n_1} \gamma_i u_i\right) d\theta\, du$$
$$\leqslant \int_{w_0} p(\theta)\left(1-\sum_{i=1}^{n_1} u_i^2\right)^{\frac{1}{2}(n-2)} \exp\left(\frac{\sqrt{t}}{\sigma}\sum_{i=1}^{n_1} \gamma_i u_i\right) d\theta\, du. \tag{41}$$

If we multiply both sides of (41) by

$$\frac{e^{-\lambda}}{(\sqrt{2}\,\sigma)^{n+n_1}\pi^{\frac{1}{2}n_1}\Gamma(\frac{1}{2}n)} p(y)\, t^{\frac{1}{2}(n+n_1-2)} \exp\left(-\frac{t}{2\sigma^2}\right),$$

and integrate over the sample space of y and t, we get, in accordance with (18),

$$\beta(\lambda) \leqslant \beta_0(\lambda).$$

Hence the theorem is proved.

APPENDIX

LEMMA. *Let $g(x) \geqslant 0$ be defined for $x \geqslant 0$ and vanish for $x > 1$, such that $g(v_1^2+\ldots+v_n^2)$ is summable. Let $f(w_1,\ldots,w_m) \geqslant 0$ be summable. In the product space of the v's and w's let R be a region such that*

$$\int_R f(w_1,\ldots,w_m)\, g(v_1^2+\ldots+v_n^2) \exp(\gamma_1 v_1+\ldots+\gamma_n v_n)\, dv\, dw = G(\gamma_1^2+\ldots+\gamma_n^2). \tag{1}$$

Let w_0 be the region defined by the inequality

$$v_1^2+\ldots+v_n^2 \geqslant k. \tag{2}$$

Let $\displaystyle\int_{R_0} f(w_1,\ldots,w_m)\, g(v_1^2+\ldots+v_n^2) \exp(\gamma_1 v_1+\ldots+\gamma_n v_n)\, dv\, dw = G_0(\gamma_1^2+\ldots+\gamma_n^2).$ (3)*

Finally, let

$$\int_R f(w_1,\ldots,w_m)\, g(v_1^2+\ldots+v_n^2)\, dv\, dw = \int_{R_0} f(w_1,\ldots,w_m)\, g(v_1^2+\ldots+v_n^2)\, dv\, dw. \tag{4}$$

Then

$$G(x) \leqslant G_0(x) \tag{5}$$

for all positive x.

* Notice that (3) is not a separate condition on R_0, but is implied by (2).

Proof. In (1) we set $\gamma_1 = x, \gamma_2 = \ldots = \gamma_n = 0$, and get

$$G(x^2) = \int_R f(w_1, \ldots, w_m)\, g(v_1^2 + \ldots + v_n^2) \exp(xv_1)\, dv\, dw.$$

This and the conditions on f and g imply that G is continuous.

Multiplying both sides of (1) by $\exp(-\Sigma\gamma_i^2)$ and integrating over the region $0 \leqslant a \leqslant \Sigma\gamma_i^2 \leqslant b$, we get

$$K\int_a^b x^{\frac{1}{2}(n-2)} e^{-x} G(x)\, dx$$
$$= \int_R f(w_1, \ldots, w_m)\, g(\Sigma v_i^2)\, dv\, dw \int_{a \leqslant \Sigma\gamma_i^2 \leqslant b} \exp(-\Sigma\gamma_i^2 + \Sigma\gamma_i v_i)\, d\gamma, \quad (6)$$

where K is some numerical constant. Applying a rotation in the space of the γ's to the inner integral in the right-hand side of (6), we obtain

$$K\int_a^b x^{\frac{1}{2}(n-2)} e^{-x} G(x)\, dx$$
$$= \int_R f(w_1, \ldots, w_m)\, g(\Sigma v_i^2)\, dv\, dw \int_{a \leqslant \Sigma x_i^2 \leqslant b} \exp\{-\Sigma x_i^2 + (\Sigma v_i^2)^{\frac{1}{2}} x_1\}\, dx$$
$$= \sum_{h=0}^{\infty} c_h I_h(R),$$

where
$$I_h(R) = \frac{1}{(2h)!}\int_R f(w_1, \ldots, w_m)\, g(\Sigma v_i^2)^h\, dv\, dw,$$
$$c_h = \int_{a \leqslant \Sigma x_i^2 \leqslant b} x_1^{2h} \exp(-\Sigma x_i^2)\, dx.$$

Similarly, we have

$$K\int_a^b x^{\frac{1}{2}(n-2)} e^{-x} G_0(x)\, dx = \sum_{h=0}^{\infty} c_h I_h(R_0).$$

An appeal to a general lemma of Neyman and Pearson,* on remembering (4), leads to the inequality

$$I_h(R) \leqslant I_h(R_0).$$

Hence
$$\int_a^b x^{\frac{1}{2}(n-2)} e^{-x}\{G(x) - G_0(x)\}\, dx \leqslant 0.$$

Since a and b are arbitrary and since the integrand is a continuous function, the latter must be $\leqslant 0$. Hence $G(x) \leqslant G_0(x)$.

* Neyman and Pearson (1936), p. 11.

REFERENCES

Doetsch, G. (1937). *Theorie und Anwendung der Laplace-Transformation*. Berlin: Julius Springer.

Fisher, R. A. (1928). "The general sampling distribution of the multiple correlation coefficient." *Proc. Roy. Soc.* A, **121**, 654–73.

Kołodziejczyk, St (1935). "On an important class of statistical hypotheses." *Biometrika*, **27**, 161–90.

Neyman, J. & Pearson, E. S. (1936). "Contributions to the theory of testing statistical hypotheses." *Statist. Res. Mem.* **1**, 1–37.

Tang, P. C. (1938). "The power function of the analysis of variance tests." *Statist. Res. Mem.* **2**, 126–49.

ON AN OPTIMUM PROPERTY OF TWO IMPORTANT STATISTICAL TESTS

By J. B. SIMAIKA, Ph.D.

P. L. Hsu (1940) has shown that for any linear hypothesis the E^2-test is the uniformly most powerful of all the tests whose power function depends on a certain function, λ, of the population parameters. Two other tests of importance, namely, those associated with the multiple correlation coefficient and Hotelling's T^2 (Hotelling, 1931), have the similar property of being uniformly more powerful than all other tests whose power functions depend on the respective functions of population parameters involved in the distributions of R^2 and T^2. It is the purpose of this paper to establish such an optimum property of these two tests. We shall consider them separately.

I. Hotelling's T^2

The general problem that calls for the T^2-test may be stated in the following way: given the elementary probability law

$$p(y_1, \ldots, y_q, s_{11}, s_{12}, \ldots, s_{qq}) = K\,|s_{ij}|^{\frac{1}{2}m-1} \exp\left\{-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}(y_i-\eta_i)(y_j-\eta_j) - \tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}s_{ij}\right\} \tag{1}$$

$$(\alpha_{ij} = \alpha_{ji}, \quad s_{ij} = s_{ji}),$$

it is required to test the hypothesis that

$$\eta_i = 0 \quad (i = 1, \ldots, q). \tag{2}$$

Hotelling's test consists in calculating

$$T^2 = \sum_{i,j=1}^{q} s^{ij} y_i y_j, \tag{3}$$

where s^{ij} denotes the general element in the matrix $\|s_{ij}\|^{-1}$, and rejecting the hypothesis if

$$T^2 \geqslant T^2_\epsilon, \tag{4}$$

where T^2_ϵ is a constant so determined that the risk of rejecting the hypothesis when it is true equals ϵ.

The distribution of T^2 derived from (1), which conforms with Fisher's (C) distribution (Fisher, 1928), was obtained independently by Hsu (1938) and Bose & Roy (1938), and may be written

$$p(T^2\,|\,\eta, \alpha) = p(T^2\,|\,\psi^2) = e^{-\psi^2} \sum_{h=0}^{\infty} \frac{(\psi^2)^h}{h!} \frac{1}{B(\frac{1}{2}q+h, \frac{1}{2}m)} (T^2)^{\frac{1}{2}q+h-1}(1+T^2)^{-\frac{1}{2}(m+q)-h}, \tag{5}$$

where

$$\psi^2 = \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}\eta_i\eta_j. \tag{6}$$

Hence the power function of the T^2-test is

$$\int_{T_\epsilon^2}^{\infty} p(T^2 \mid \psi^2)\, d(T^2), \tag{7}$$

which depends only on the function ψ^2 of the η's and α's. Our first theorem asserts that the T^2-test is uniformly more powerful than any other test whose power function is a function of ψ^2 alone.

THEOREM I. *Let w_0 (of size ϵ) be the critical region defined by the inequality* (4), *and w be any other critical region whose size is ϵ and whose power function is a function of ψ^2. Let $\beta(\psi^2)$ and $\beta_0(\psi^2)$ be the power functions of w and w_0 respectively. Then*

$$\beta(\psi^2) \leqslant \beta_0(\psi^2). \tag{8}$$

Proof. Let us first find a necessary and sufficient condition that w should have the properties described in Theorem I. We have, from (1),

$$p(y, s) = Ke^{-\psi^2} \mid s_{ij} \mid^{\frac{1}{2}m-1} \exp\left\{-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}(s_{ij}+y_i y_j) + \sum_{i,j=1}^{q} \alpha_{ij}\eta_i y_j\right\}. \tag{9}$$

Hence, on setting $\quad u_{ij} = s_{ij}+y_i y_j \quad (i,j = 1, \ldots, q),$ (10)

and

$$\zeta_i = \sum_{j=1}^{q} \alpha_{ij}\eta_j \quad (i = 1, \ldots, q), \tag{11}$$

we have

$$p(y,u) = Ke^{-\psi^2} \mid u_{ij}-y_i y_j \mid^{\frac{1}{2}m-1} \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}u_{ij} + \sum_{i=1}^{q} \zeta_i y_i\right) \tag{12}$$

and

$$\psi^2 = \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha^{ij}\zeta_i\zeta_j, \tag{13}$$

where α^{ij} denotes the general element of the matrix $\|\alpha_{ij}\|^{-1}$.

If w is of size ϵ and has a power function depending only on ψ^2, then

$$K\int_w |u_{ij} - y_i y_j|^{\frac{1}{2}m-1} \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}u_{ij}\right) dy\, du = \epsilon \tag{14}$$

and

$$K\int_w \mid u_{ij}-y_i y_j \mid^{\frac{1}{2}m-1} \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}u'_{ij} + \sum_{i=1}^{q} \zeta_i y_i\right) dy\, du = e^{\psi^2}\beta(\psi^2) = F(\psi^2), \text{ say.} \tag{15}$$

It follows from (15) that, on expanding the left-hand side into a power series in the ζ's, we must have

$$\int_w \mid u_{ij}-y_i y_j \mid^{\frac{1}{2}m-1} \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}u_{ij}\right)\left(\sum_{i=1}^{q} \zeta_i y_i\right)^h dy\, du = 0 \text{ for odd } h, \tag{16}$$

$$\frac{K}{(2h)!}\int_w \mid u_{ij}-y_i y_j \mid^{\frac{1}{2}m-1} \exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij}u_{ij}\right)\left(\sum_{i=1}^{q} \zeta_i y_i\right)^{2h} dy\, du = \frac{1}{h!}a_h(\psi^2)^h \quad (h = 1, 2, 3, \ldots), \tag{17}$$

where the a_h are numbers depending only on the region w chosen.

On the other hand, since the integral of (12) over the sample space W is unity, we have

$$K\int_W |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}+\sum_{i=1}^{q}\zeta_iy_i\right)dy\,du = e^{\psi^2}, \tag{18}$$

whence

$$K\int_W |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)dy\,du = 1, \tag{19}$$

$$\frac{K}{(2h)!}\int_W |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h}dy\,du$$
$$=\frac{1}{h!}(\psi^2)^h \quad (h=1,2,3,\ldots). \tag{20}$$

Combining equations (14) and (19), (17) and (20), we obtain

$$\int_w |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)dy\,du$$
$$=\epsilon\int_W |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)dy\,du, \tag{21}$$

$$\int_w |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h}dy\,du$$
$$=a_h\int_W |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h}dy\,du$$
$$(h=1,2,3,\ldots). \tag{22}$$

The sample space W is the product space $W(u)\times W(y\mid u)$, where $W(u)$ is the sample space of the u's and $W(y\mid u)$ is formed of the possible positions of the point $(y_1,\ldots,y_q)$ for given values of the u's. Similarly $w=W(u)\times w(y\mid u)$. If we evaluate the integrals in (21), (16) and (22) as repeated integrals, we obtain

$$\int_{W(u)}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)$$
$$\times\left[\int_{w(y|u)}|u_{ij}-y_iy_j|^{\frac{1}{2}m-1}dy-\epsilon\int_{W(y|u)}|u_{ij}-y_iy_j|^{\frac{1}{2}m-1}dy\right]du=0, \tag{23}$$

$$\int_{W(u)}\exp\left(-\tfrac{1}{2}\sum_{i=1}^{q}\alpha_{ij}u_{ij}\right)\left[\int_{w(y|u)}|u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{h}dy\right]du=0 \text{ for odd } h, \tag{24}$$

$$\int_{W(u)}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}u_{ij}\right)\left[\int_{w(y|u)}|u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h}dy-a_h\right.$$
$$\left.\times\int_{W(y|u)}|u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h}dy\right]du=0\quad(h=1,2,3,\ldots). \tag{25}$$

Since equations (23), (24) and (25) must hold true for all admissible sets of values

of the α_{ij}, so, according to the lemma proved in the Appendix, the functions within the square brackets in these equations must vanish identically. Hence

$$\int_{w(y|u)} |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\,dy = \epsilon\int_{W(y|u)} |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\,dy, \tag{26}$$

$$\int_{w(y|u)} |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^h dy = 0 \text{ for odd } h, \tag{27}$$

$$\int_{w(y/u)} |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h} dy = a_h\int_{W(y/u)} |u_{ij}-y_iy_j|^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\zeta_iy_i\right)^{2h} dy$$
$$(h = 1, 2, 3, \ldots). \tag{28}$$

In order to simplify the above equations we notice that the matrix $\|u_{ij}\|$, being positive definite, can be thrown into the form $\mathbf{CC}'$, where $\mathbf{C}$ is a non-singular real matrix. Using the transformation

$$\|y_1, \ldots, y_q\| = \|x_1, \ldots, x_q\|\,\mathbf{C}', \tag{29}$$

we get that

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1} dx = \epsilon\int_{W(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1} dx, \tag{30}$$

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\xi_ix_i\right)^h dx = 0 \text{ for odd } h, \tag{31}$$

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\xi_ix_i\right)^{2h} dx$$
$$= a_h\int_{W(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\xi_ix_i\right)^{2h} dx \quad (h = 1, 2, 3, \ldots), \tag{32}$$

where

$$\|\xi_1, \ldots, \xi_q\| = \|\zeta_1, \ldots, \zeta_q\|\,\mathbf{C}. \tag{33}$$

Now $W(x\,|\,u)$ is the region S (independent of the u's) defined by the inequality

$$\sum_{i=1}^{q}x_i^2 \leqslant 1. \tag{34}$$

Hence the integral in the right-hand side of (30) is a numerical constant, say b, and a rotation in the space of the x's enables the integral in the right-hand side of (32) to be written as

$$\left(\sum_{i=1}^{q}\xi_i^2\right)^h\int_S x_1^{2h}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1} dx. \tag{35}$$

Hence we obtain the following equivalents of equations (30), (31) and (32):

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1} dx = b\epsilon, \tag{36}$$

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\xi_ix_i\right)^h dx = 0 \text{ for odd } h, \tag{37}$$

$$\int_{w(x|u)}\left(1-\sum_{i=1}^{q}x_i^2\right)^{\frac{1}{2}m-1}\left(\sum_{i=1}^{q}\xi_ix_i\right)^{2h} dx = b_h\left(\sum_{i=1}^{q}\xi_i^2\right)^h \quad (h = 1, 2, 3, \ldots), \tag{38}$$

where the b_h are numbers depending only on the choice of the region w.

The set of equations (36), (37) and (38) give the necessary and sufficient condition that the critical region w should have the properties described in Theorem I. Further, equations (37) and (38) may be combined into the following:

$$\int_{w(x|u)} \left(1-\sum_{i=1}^{q} x_i^2\right)^{\frac{1}{2}m-1} \exp\left(\sum_{i=1}^{q} \xi_i x_i\right) dx = G\left(\sum_{i=1}^{q} \xi_i^2\right). \tag{39}$$

Now according to (29) we have

$$\sum_{i=1}^{q} x_i^2 = \sum_{i,j=1}^{q} u^{ij} y_i y_j = T^2/(1+T^2), \tag{40}$$

where u^{ij} denotes the general element of the matrix $\|u_{ij}\|$. Hence w_0 is the region defined by the inequality

$$\sum_{i=1}^{q} x_i^2 \geqslant T_\epsilon^2/(1+T_\epsilon^2). \tag{41}$$

Since w_0 is of size ϵ, we must have the same equation as (36) when w is replaced by w_0 therein. Hence

$$\int_{w(x|u)} \left(1-\sum_{i=1}^{q} x_i^2\right)^{\frac{1}{2}m-1} dx = \int_{w_0} \left(1-\sum_{i=1}^{q} x_i^2\right)^{\frac{1}{2}m-1} dx. \tag{42}$$

Letting $$\int_{w_0} \left(1-\sum_{i=1}^{q} x_i^2\right)^{\frac{1}{2}m-1} \exp\left(\sum_{i=1}^{q} \xi_i x_i\right) dx = G_0\left(\sum_{i=1}^{q} \xi_i^2\right), \tag{43}$$

we deduce with the help of (41), (42) and the lemma proved by P. L. Hsu in the Appendix of his paper (1940) that

$$G\left(\sum_{i=1}^{q} \xi_i^2\right) \leqslant G_0\left(\sum_{i=1}^{q} \xi_i^2\right). \tag{44}$$

Applying the transformation reciprocal to (29) to the integrals in (39) and (43), we get

$$\int_{w(y|u)} |u_{ij}-y_i y_j|^{\frac{1}{2}m-1} \exp\left(\sum_{i=1}^{q} \zeta_i y_i\right) dy \leqslant \int_{w_0(y|u)} |u_{ij}-y_i y_j|^{\frac{1}{2}m-1} \exp\left(\sum_{i=1}^{q} \zeta_i y_i\right) dy. \tag{45}$$

Hence on multiplying both sides of (45) by $K \exp(-\psi^2) \exp\left(-\frac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} u_{ij}\right)$ and integrating over $W(u)$ and remembering (15), we have the inequality (8):

$$\beta(\psi^2) \leqslant \beta_0(\psi^2),$$

which was to be proved.

II. MULTIPLE CORRELATION COEFFICIENT

In this connection the basic elementary probability law is taken to be

$$p(z, y_1, \ldots, y_q, x_{11}, x_{12}, \ldots, x_{qq}) = K(1-\rho^2)^{\frac{1}{2}n} \begin{vmatrix} z & y_1 & \cdots & y_q \\ y_1 & x_{11} & \cdots & x_{1q} \\ \cdots & \cdots & \cdots & \cdots \\ y_q & x_{q1} & \cdots & x_{qq} \end{vmatrix}^{\frac{1}{2}(n-q-2)}$$

$$\times \exp\left(-\tfrac{1}{2}\gamma z - \sum_{i=1}^{q} \beta_i y_i - \tfrac{1}{2} \sum_{i,j=1}^{q} \alpha_{ij} x_{ij}\right) \quad (1)$$

$$(\alpha_{ij} = \alpha_{ji}, \quad x_{ij} = x_{ji}),$$

where
$$\rho^2 = \frac{1}{\gamma} \sum_{i,j=1}^{q} \alpha^{ij} \beta_i \beta_j \quad (2)$$

is the square of the multiple correlation coefficient of the population. We have

$$\begin{vmatrix} z & y_1 & \cdots & y_q \\ y_1 & x_{11} & \cdots & x_{1q} \\ \cdots & \cdots & \cdots & \cdots \\ y_q & x_{q1} & \cdots & x_{qq} \end{vmatrix} = |x_{ij}| \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right) \quad (3)$$

and that the square of the multiple correlation coefficient of the sample is

$$R^2 = \frac{1}{z} \sum_{i,j=1}^{q} x^{ij} y_i y_j. \quad (4)$$

The hypothesis to be tested is that

$$\beta_i = 0 \quad (i = 1, \ldots, q). \quad (5)$$

THEOREM II. *The basic elementary probability law and the hypothesis under test being given by* (1) *and* (5), *let* w_0 *be the critical region of size* ϵ *defined by the inequality*

$$R^2 \geqslant R_\epsilon^2 \quad (6)$$

and w *be any other critical region whose size is* ϵ *and whose power function depends only on* ρ^2. *Let* $\beta(\rho^2)$ *and* $\beta_0(\rho^2)$ *be the power functions of* w *and* w_0 *respectively. Then*

$$\beta(\rho^2) \leqslant \beta_0(\rho^2). \quad (7)$$

Proof. Suppose that w has the properties described in Theorem II. Then

$$K\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \exp\left(-\tfrac{1}{2}\gamma z - \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} x_{ij}\right) dz\, dy\, dx = \epsilon, \quad (8)$$

$$K\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)}$$

$$\times \exp\left(-\tfrac{1}{2}\gamma z - \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} x_{ij} - \sum_{i=1}^{q} \beta_i y_i\right) dz\, dy\, dx$$

$$= (1-\rho^2)^{-\frac{1}{2}n} \beta(\rho^2) = F(\rho^2), \text{ say.} \quad (9)$$

Hence, on developing the left-hand side of (9) into a power series in the β's, we have

$$\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right)\left(\sum_{i=1}^{q}\beta_i y_i\right)^h dz\,dy\,dx = 0 \text{ for odd } h, \tag{10}$$

$$\frac{K}{(2h)!}\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right)\left(\sum_{i=1}^{q}\beta_i y_i\right)^{2h} dz\,dy\,dx = \frac{\Gamma(\frac{1}{2}n+h)}{h!\,\Gamma(\frac{1}{2}n)}a_h(\rho^2)^h \quad (h=1,2,3,\ldots), \tag{11}$$

where the a_h are numbers depending only on the choice of the region w.

On the other hand, since the integral of (1) over the whole sample space W is unity, we have

$$K\int_W |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}-\sum_{i=1}^{q}\beta_i y_i\right) dz\,dy\,dx = (1-\rho^2)^{-\frac{1}{2}n}, \tag{12}$$

whence

$$K\int_W |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right) dz\,dy\,dx = 1, \tag{13}$$

$$\frac{K}{(2h)!}\int_W |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right)\left(\sum_{i=1}^{q}\beta_i y_i\right)^{2h} dz\,dy\,dx = \frac{\Gamma(\frac{1}{2}n+h)}{h!\,\Gamma(\frac{1}{2}n)}(\rho^2)^h \quad (h=1,2,3,\ldots). \tag{14}$$

Combining equations (8) and (13), (11) and (14), we obtain

$$\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right) dz\,dy\,dx = \epsilon\int_W (\ldots)\,dz\,dy\,dx, \tag{15}$$

$$\int_w |x_{ij}|^{\frac{1}{2}(n-q-2)}\left(z-\sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \times \exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right)\left(\sum_{i=1}^{q}\beta_i y_i\right)^{2h} dz\,dy\,dx = a_h\int_W (\ldots)\,dz\,dy\,dx \quad (h=1,2,3,\ldots), \tag{16}$$

where the unwritten integrands in the right-hand sides are the same as those in the left-hand sides.

As before we argue that $W = W(z,x) \times W(y \mid z,x)$, $w = W(z,x) \times w(y \mid z,x)$ and evaluate the integrals in (15), (10) and (16) as repeated integrals. It follows that

$$\int_{W(z,x)} |x_{ij}|^{\frac{1}{2}(n-q-2)} \exp\left(-\tfrac{1}{2}\gamma z - \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} x_{ij}\right) \left[\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} dy \right.$$
$$\left. - \epsilon \int_{W(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} dy \right] dz\, dx = 0, \qquad (17)$$

$$\int_{W(z,x)} |x_{ij}|^{\frac{1}{2}(n-q-2)} \exp\left(-\tfrac{1}{2}\gamma z - \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} x_{ij}\right)$$
$$\left[\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \beta_i y_i\right)^h dy\right] dz\, dx = 0 \text{ for odd } h, \qquad (18)$$

$$\int_{W(z,x)} |x_{ij}|^{\frac{1}{2}(n-q-2)} \exp\left(-\tfrac{1}{2}\gamma z - \tfrac{1}{2}\sum_{i,j=1}^{q} \alpha_{ij} x_{ij}\right)$$
$$\left[\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \beta_i y_i\right)^{2h} dy\right.$$
$$\left. - a_h \int_{W(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \beta_i y_i\right)^{2h}\right] dy\, dz\, dx = 0 \quad (h = 1, 2, 3, \ldots). \qquad (19)$$

According to the lemma proved in the Appendix, the functions within the square brackets in the above equations must vanish identically. Hence

$$\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} dy = \epsilon \int_{W(y|z,x)} (\ldots)\, dy, \qquad (20)$$

$$\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \beta_i y_i\right)^h dy = 0 \text{ for odd } h, \qquad (21)$$

$$\int_{w(y|z,x)} \left(z - \sum_{i,j=1}^{q} x^{ij} y_i y_j\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \beta_i y_i\right)^{2h} dy$$
$$= a_h \int_{w(y|z,x)} (\ldots)\, dy \quad (h = 1, 2, 3, \ldots). \qquad (22)$$

In order to simplify the above equations we notice that, since the matrix $\|x_{ij}\|$ is positive definite, it can be thrown into the form $\mathbf{CC}'$, where $\mathbf{C}$ is a non-singular real matrix. Using the transformation

$$\|y_1, \ldots, y_q\| = z^{\frac{1}{2}} \|t_1, \ldots, t_q\| \mathbf{C}', \qquad (23)$$

we obtain
$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} dt = \epsilon \int_{W(t|z,x)} (\ldots)\, dt, \tag{24}$$

$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \tau_i t_i\right)^{h} dt = 0 \text{ for odd } h, \tag{25}$$

$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \tau_i t_i\right)^{2h} dt = a_h \int_{W(t|z,x)} (\ldots)\, dt \quad (h = 1, 2, 3, \ldots), \tag{26}$$

where
$$\|\tau_1, \ldots, \tau_q\| = z^{\frac{1}{2}} \|\beta_1, \ldots, \beta_q\| \mathbf{C}.$$

Now $W(t \mid z, x)$ is the region S (independent of z and the x's) defined by the inequality
$$\sum_{i=1}^{q} t_i^2 \leqslant 1. \tag{27}$$

Hence the integral on the right-hand side of (24) is a numerical constant, say b, and a rotation in the space of the t's enables the integral on the right-hand side of (26) to be written as
$$\left(\sum_{i=1}^{q} \tau_i^2\right)^{h} \int_S \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} t_1^{2h}\, dt. \tag{28}$$

Hence we have the following equivalents of (24), (25) and (26):
$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} dt = b\epsilon, \tag{29}$$

$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \tau_i t_i\right)^{h} dt = 0 \text{ for odd } h, \tag{30}$$

$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} \left(\sum_{i=1}^{q} \tau_i t_i\right)^{2h} dt = b_h \left(\sum_{i=1}^{q} \tau_i^2\right)^{h} \quad (h = 1, 2, 3, \ldots), \tag{31}$$

where the b_h are numbers depending only on the choice of w.

Equations (30) and (31) may be combined into the following one:
$$\int_{w(t|z,x)} \left(1-\sum_{i=1}^{q} t_i^2\right)^{\frac{1}{2}(n-q-2)} \exp\left(-\sum_{i=1}^{q} \tau_i t_i\right) dt = G\left(\sum_{i=1}^{q} \tau_i^2\right). \tag{32}$$

Now, by (4) and (23), we have
$$R^2 = \sum_{i=1}^{q} t_i^2; \tag{33}$$

consequently the region w_0 is defined by the inequality
$$\sum_{i=1}^{q} t_i^2 \geqslant R_\epsilon^2. \tag{34}$$

Since w_0 is of size ϵ, we must have the same equation (29) when $w(t\,|\,z,x)$ is replaced by w_0 therein. Hence

$$\int_{w(t|z,x)}\left(1-\sum_{i=1}^{q}t_i^2\right)^{\frac{1}{2}(n-q-2)}dt=\int_{w_0}\left(1-\sum_{i=1}^{q}t_i^2\right)^{\frac{1}{2}(n-q-2)}dt. \tag{35}$$

On setting $$\int_{w_0}\left(1-\sum_{i=1}^{q}t_i^2\right)^{\frac{1}{2}(n-q-2)}\exp\left(-\sum_{i=1}^{q}\tau_i t_i\right)dt=G_0\left(\sum_{i=1}^{q}\tau_i^2\right), \tag{36}$$

we infer, with the help of (34), (35) and the lemma proved by P. L. Hsu in the Appendix of his paper (1940), that

$$G\left(\sum_{i=1}^{q}t_i^2\right)\leqslant G_0\left(\sum_{i=1}^{q}t_i^2\right). \tag{37}$$

Hence, using the transformation reciprocal to (23) to the integrals in (32) and (36), we have

$$\int_{w(y|z,x)}\left(z-\sum_{i,j=1}^{q}x^{ij}y_iy_j\right)^{\frac{1}{2}(n-q-2)}\exp\left(-\sum_{i=1}^{q}\beta_iy_i\right)dy\leqslant\int_{w_0(y|z,x)}(\ldots)\,dy. \tag{38}$$

Multiplying both sides of (38) by

$$K(1-\rho^2)^{\frac{1}{2}n}\,|\,x_{ij}\,|^{\frac{1}{2}(n-q-2)}\exp\left(-\tfrac{1}{2}\gamma z-\tfrac{1}{2}\sum_{i,j=1}^{q}\alpha_{ij}x_{ij}\right)$$

and integrating over the space $W(z,x)$ and remembering (9), we obtain

$$\beta(\rho^2)\leqslant\beta_0(\rho^2).$$

Therefore Theorem II is proved.

I am gratefully indebted to Dr P. L. Hsu for putting this problem before me and for his helpful suggestions both in the course of my research and in preparing this paper for publication.

APPENDIX

LEMMA. *Let $E(x)$ be the set of points $(x_{11}, x_{12}, \ldots, x_{qq})$ for which the symmetric matrix $\|x_{ij}\|$ is positive definite. Then*

$$\int_{E(x)}|\,x_{ij}\phi(x)\,|\exp\left(-\sum_{i=1}^{q}x_{ij}\right)dx<\infty\quad(i,j=1,\ldots,q) \tag{1}$$

and

$$\int_{E(x)}\phi(x)\exp\left(-\sum_{i=1}^{q}\alpha_{ij}x_{ij}\right)dx=0 \textit{ throughout } E(\alpha)\quad(x_{ij}=x_{ji},\ \alpha_{ij}=\alpha_{ji}) \tag{2}$$

imply that $\phi(x)=0$ *almost everywhere in* $E(x)$. (3)

Proof. Suppose that both (1) and (2) are true. Since the matrix $\|\delta_{ij}+\theta_{ij}\|$, where $\delta_{ii}=1$ and $\delta_{ij}=0$ $(i\neq j)$, $\theta_{ij}=\theta_{ji}$, is positive definite for all sufficiently small real θ's; so, by (2),

$$\int_{E(x)} \phi(x) \exp\left(-\sum_{i=1}^{q} x_{ii}\right) \exp\left(-\sum_{i=1}^{q} \theta_{ij} x_{ij}\right) dx = 0 \tag{4}$$

for all sufficiently small real θ's. By (1) the left-hand side of (4) is an analytic function of each of the θ's in the neighbourhood of the imaginary axis. By analytic continuation (4) must remain true for all complex θ's with sufficiently small real parts. In particular,

$$\int_{E(x)} \phi(x) \exp\left(-\sum_{i=1}^{q} x_{ii}\right) \exp\left(\sqrt{-1} \sum_{i,j=1}^{q} t_{ij} x_{ij}\right) dx = 0 \quad (t_{ij}=t_{ji}) \tag{5}$$

for all real values of t's. Hence, by the well-known property of the Fourier transform,

$$\phi(x) \exp\left(-\sum_{i=1}^{q} x_{ii}\right) = 0 \text{ almost everywhere in } E(x),$$

which implies (3).

REFERENCES

Bose, R. C. & Roy, S. N. (1938). "The distribution of the 'Studentized' D^2-statistic." *Proceedings of the Indian Statistical Conference, Calcutta*, pp. 19–38.

Fisher, R. A. (1928). "The general sampling distribution of the multiple correlation coefficient." *Proc. Roy. Soc.* A, **121**, 654–73.

Hotelling, H. (1931). "The generalization of Student's ratio." *Ann. Math. Statist.* **2**, 359–78.

Hsu, P. L. (1938). "Notes on Hotelling's generalized T." *Ann. Math. Statist.* **9**, 231–43.

—— (1940). "Analysis of Variance from the power function standpoint." *Biometrika*, **32**, 62–9.

MISCELLANEA

(i) A recurrence relation for the semi-invariants of Pearson curves

By M. G. KENDALL

The Pearson curves are defined by the differential equation

$$dy = \frac{y(a+x)\,dx}{b_0+b_1x+b_2x^2}.$$

Multiplying by $e^{tx}(b_0+b_1x+b_2x^2)$ and integrating over the range of the distribution, we have

$$\int e^{tx}y(a+x)\,dx = \int (b_0+b_1x+b_2x^2)\,e^{tx}\,dy$$
$$= [(b_0+b_1x+b_2x^2)\,e^{tx}y] - \int y\,dx\,e^{tx}\{b_1+2b_2x+t(b_0+b_1x+b_2x^2)\}.$$

At the extremes of the distribution we may suppose the expression in square brackets on the right to vanish and hence

$$\int e^{tx}y\{a+b_1+b_0t+(1+2b_2+b_1t)\,x+b_2tx^2\}\,dx = 0. \qquad \ldots\ldots(1)$$

The moment generating function of the distribution, $M(t)$, is given by

$$M(t) = \int e^{tx}y\,dx,$$

and hence $\dfrac{dM}{dt} = \displaystyle\int e^{tx}xy\,dx$, etc. Thus from (1)

$$b_2t\frac{d^2M}{dt^2}+(1+2b_2+b_1t)\frac{dM}{dt}+(a+b_1+b_0t)\,M = 0, \qquad \ldots\ldots(2)$$

a linear differential equation of the second order, which may also be regarded as defining the Pearsonian system.

Incidentally, it would be interesting from the theoretical view-point to consider classes of frequency distributions defined by differential equations in their moment or semi-invariant generating functions.

So far as I know there is no solution of (2) in ordinary functions which would permit of the explicit expression of the co-efficient of t^r in $M(t)$; but from a consideration of the co-efficient of t^r in (2) we have

$$\{1+(r+2)\,b_2\}\,\mu'_{r+1}+\{a+(r+1)\,b_1\}\,\mu'_r+rb_0\mu'_{r-1} = 0, \qquad \ldots\ldots(3)$$

the well-known recurrence relation between the moments of Pearson curves.

Some simplification of this expression is possible by the choice of a particular origin in certain cases. If the roots of

$$b_0+b_1x+b_2x^2 = 0$$

are real, it is possible by a real linear transformation to transform the equation defining the Pearson curves to one which does not involve b_0. With the origin defined by this transformation, we have

$$\{1+(r+2)\,b_2\}\,\mu'_{r+1}+\{a+(r+1)\,b_1\}\,\mu'_r = 0,$$

giving

$$\mu'_r = (-1)^r\frac{(a+rb_1)\,(a+\overline{r-1}b_1)\ldots(a+b_1)}{(1+\overline{r+1}b_2)\,(1+rb_2)\ldots(1+2b_2)}. \qquad \ldots\ldots(4)$$

Putting $K = \log M$ in (2), we have for the semi-invariant generating function

$$b_2 t\left\{\frac{d^2K}{dt^2}+\left(\frac{dK}{dt}\right)^2\right\}+(1+2b_2+b_1 t)\frac{dK}{dt}+(a+b_1+b_0 t) = 0. \quad \text{......(5)}$$

This is not linear, and it appears therefore that there is no simple recurrence relation among the semi-invariants as among the moments. The equation is similar in character to that known as Riccati's and the usual way of solving it would be to return to the linear equation (2) from which it was derived.

Taking an origin at the mean ($\kappa_1 = 0$) and considering the co-efficient of t^r in (5), we have

$$\frac{b_2\kappa_{r+1}}{(r-1)!}+b_2\left\{\frac{\kappa_2}{1!}\frac{\kappa_{r-1}}{(r-2)!}+\frac{\kappa_3}{2!}\frac{\kappa_{r-2}}{(r-3)!}+\ldots+\frac{\kappa_{r-1}}{(r-2)!}\frac{\kappa_2}{1!}\right\}+(1+2b_2)\frac{\kappa_{r+1}}{r!}+b_1\frac{\kappa_r}{(r-1)!} = 0,$$

or

$$\{1+(r+2)\,b_2\}\,\kappa_{r+1}+rb_1\kappa_r+rb_2\left\{\binom{r-1}{1}\kappa_2\kappa_{r-1}\right.$$

$$\left.+\binom{r-1}{2}\kappa_3\kappa_{r-2}+\ldots+\binom{r-1}{j}\kappa_{j+1}\kappa_{r-j}+\ldots+\binom{r-1}{1}\kappa_{r-1}\kappa_2\right\} = 0, \quad \text{......(6)}$$

with the initial relation $\qquad \kappa_2 = -b_0/(1+3b_2)$.

Equation (6) seems to be as simple a recurrence relation as we can expect for the expression of a semi-invariant in terms of those of lower order.

(ii) A comparison of annual and biennial inflorescences of *Daucus carota* (wild carrot)

By WILLIAM DOWELL BATEN
Michigan State College

Introduction

In 1932 seeds from Michigan and Indiana were gathered from *Daucus carota* for the purpose of studying environmental effects on the numbers of pedicels and bracts per inflorescence from plants grown from Michigan and Indiana seeds. In 1933 these seeds were planted in the botanical gardens of the University of Michigan in the green house and later planted outside. In 1933, 44 % of the plants bloomed; in 1934, 17 % of those that did not bloom the first season survived the winter and bloomed. Results of this study were published by the present writer (1934) in an article entitled "A statistical study of *Daucus carota*", in which the numbers of pedicels and bracts on annual and biennial inflorescences coming from these seeds were compared. At the end of the article K. Pearson pointed out that since the seeds were taken from many plants in the wild, some of the seeds might have come from flowers blooming the first season and others from those blooming the second season, that one did not know how many annual and biennial seeds came from the two states and that the comparisons might not be the same if this was considered.

To overcome this just criticism seeds were taken from one plant near Ann Arbor, Michigan in 1936. These were planted in 1937 in the greenhouse at Michigan State College and later planted outside in rows 3 ft. apart and 3 ft. apart in the rows. During the latter part of the summer, counts were made on plants blooming the first year of the number of branches, the number of inflorescences, and the number of primary pedicels and bract per inflorescence on the stem and first eight branches below the stem terminal cluster. During the summer

of 1938 similar counts were made on the plants blooming the second year. The object of this article is to compare the counts pertaining to the annual and biennial inflorescences.

During the first flowering season 60 % of the plants bloomed. At the end of this season the plants which did not bloom appeared to be in good condition for the coming winter. In 1938 biennial flowers appeared much earlier than the annual flowers in 1937. Counts of the annuals were made in 1937 in August and September; counts of the biennials were made in July and the first part of August.

The terminal inflorescences on the stem will be designated by T, the first branch terminal inflorescence by A, the first non-terminal inflorescence on the first branch by A_1, etc. Branches are considered in descending order below the stem terminal. According to these notations, D_3 represents the third non-terminal inflorescence on the fourth branch. Umbels in this article will always mean primary umbels and pedicels or rays will always mean primary pedicels or rays.

Size of annual and biennial plants

The following averages pertain to the number of branches and inflorescences (including buds) of annual and biennial plants.

Parts	Annuals (1937)	Biennials (1938)
Average no. of branches	16·3	20·3
Average no. of inflorescences	122·7	282·5

These averages indicate that the biennial plants were much larger as to number of branches and inflorescences than the annual plants. The second year herbs were considerably taller than those blooming the first season.

In 1938 most of the branches used in making the counts had four umbels, whose parts could be enumerated; in 1937 very few of these had four umbels which were mature enough to use. Very few of the first branches belonging to 1937 plants produced more than two non-terminal umbels; a good percentage of corresponding branches of 1938 plants possessed more than two. Counts were made on seventy-seven plants during the first season and on seventy-six during the second. In the second summer there were several plants with more than 500 inflorescences and one with 796; the largest in 1937 had 183.

In 1937 there was 37·7 % of the herbs with at least eight branches; in 1938 there was 72·4 % with at least eight similar branches. In the first flowering season 46·8 % of the plants had at least six branches; during the second season 90·8 % had at least six branches. There were 74·0 % of the annual plants with at least four branches and 97·4 % of the biennials with at least four. These figures show that the biennial plants were more completely filled out than the annuals.

Size of inflorescences

Table 1 contains averages pertaining to the number of bracts per umbel on the stem and the first three branches. On the average the number of bracts on the stem and branch terminal clusters of biennials are significantly larger than similar annual clusters. The average size (in number of bracts) of stem umbels for annuals was 10·9 bracts; that for biennials was 11·9 bracts. The average number of bracts per branch terminal was less than 10·3 bracts during 1937 and greater than 11·4 bracts during 1938. These figures and figures pertaining to the first eight branches indicate that the averages of the number of bracts on the majority of the biennial clusters are significantly larger than similar averages with respect to annual clusters.

Table 1. *Averages and standard deviations pertaining to the number of bracts per umbel on the stem and first three branches*

1937

	T	A	A_1	A_2	A_3	B	B_1	B_2	B_3	C	C_1	C_2	C_3
Number	77	77	55	18	—	76	59	49	—	68	52	39	—
Average	10·9	9·9	9·4	8·6	—	10·4	9·5	9·5	—	10·3	9·4	9·4	—
Standard deviation	1·22	1·39	1·22	1·01	—	1·65	1·33	1·30	—	1·42	1·21	1·27	—

1938

	T	A	A_1	A_2	A_3	B	B_1	B_2	B_3	C	C_1	C_2	C_3
Number	76	76	33	27	7	75	30	29	15	71	30	25	21
Average	11·9	11·4	9·8	10·3	11·0	11·6	10·1	10·7	11·0	11·9	10·2	10·9	11·2
Standard deviation	1·35	1·20	1·43	1·53	1·77	1·20	1·41	1·44	1·15	1·10	1·32	1·45	1·23

Biennial stem terminals had on the average significantly more pedicels than stem annuals; these averages are:

Annuals	Biennials
55·6 pedicels	67·8 pedicels

Branch terminals of biennials have on the average significantly more rays than similar ones on annuals. Fig. 1 allows the eye to see at once how these averages compare; the heights of the bars on the left represent the averages for the annuals. The bars on the left are shorter in every case.

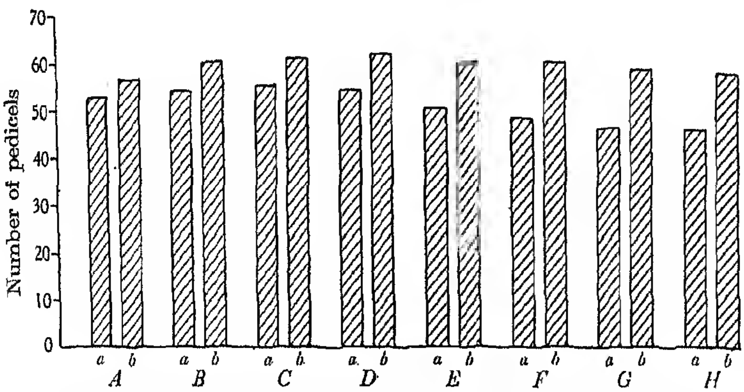


Fig. 1. Averages pertaining to number of pedicels per branch terminal umbel for annuals and biennials. *a*, annuals; *b*, biennials.

Many of the averages of the numbers of pedicels on non-terminal biennial clusters are significantly larger than those belonging to corresponding annual clusters. The above indicates that biennial inflorescences (in number of bracts and pedicels) are significantly larger

than corresponding annual inflorescences, showing again that *Daucus carota* herbs blooming the second season are on the average much larger than those blooming the first.

On examining average number of pedicels per umbel it is found that branch non-terminal umbels have on the average a smaller number of pedicels than the corresponding branch terminals; for example, A_1 and A_2 are significantly less (in number of rays) than A. This was true for the other branches. The 1938 averages for C, C_1, C_2 and C_3 are as follows:

C	C_1	C_2	C_3
61·7 rays	45·3 rays	49·3 rays	51·1 rays

Similar figures were found for the other branches. The averages pertaining to pedicels are shown in Table 2.

Table 2. *Averages and standard deviations pertaining to the number of pedicels per umbel on the stem and first three branches*

1937

	T	A	A_1	A_2	A_3	B	B_1	B_2	B_3	C	C_1	C_2	C_3
Number	77	77	55	18	—	76	59	49	—	68	52	39	—
Average	55·6	53·2	44·5	40·2	—	54·3	46·0	46·5	—	55·6	46·8	45·6	—
Standard deviation	12·65	10·98	7·40	6·96	—	11·32	8·81	9·37	—	12·20	8·70	8·18	—

1938

	T	A	A_1	A_2	A_3	B	B_1	B_2	B_3	C	C_1	C_2	C_3
Number	76	76	33	27	7	75	30	29	15	71	30	25	21
Average	64·26	57·01	45·06	48·07	48·43	60·77	47·23	49·45	47·13	61·69	45·33	49·32	51·14
Standard deviation	13·10	9·78	9·98	5·27	8·79	10·58	9·93	10·31	10·60	11·12	11·62	10·07	9·07

Correlation

The Pearson linear correlation coefficient between the number of bracts and the number of rays for various umbels for annual and biennial umbels are as follows:

Umbel ...	T	A	B	C
Annuals	0·406	0·410	0·496	0·571
Biennials	0·655	0·612	0·590	0·549

All of these coefficients are significantly different from zero, showing that there is a definite relation between the number of bracts and the number of rays. There are no significant differences between the correlation coefficients pertaining to annuals and biennials except that for T which is barely significant at the 5 % level. These values suggest that the size of the plant and season do not effect the relation between the number of bracts and rays per umbel. Similar figures were found in other investigations of this species (Baten, 1934). The position of the umbels on the plant also does not affect the relation between bracts and rays.

The coefficients of correlation between the number of bracts on T and on the other clusters pertaining to annuals and biennials are about the same and are significant, indicating a real association between bracts on stem and branch terminals and similarly for rays. The values of r_{TB} are:

	Bracts		Rays	
	Annuals	Biennials	Annuals	Biennials
r_{TB}	0·644	0·596	0·849	0·741

Size of herb and season have no effect on the relation between bracts and rays on T and on A, B and C.

The amounts of dependence of the number of bracts on branch non-terminals have on the number of bracts of terminals for the first three branches were obtained by the correlation coefficients between these respective numbers. There are no significant differences between these coefficients, suggesting that the size of plants and seasons do not affect the relation between the number of bracts and rays on branch terminals and branch first non-terminals. This also was true for rays.

The following figures are the coefficients of correlation between the number of bracts and rays on first and second branch terminal inflorescences.

Description	Bracts		Rays	
	Annuals	Biennials	Annuals	Biennials
r_{AB} (interclass)	0·748	0·734	0·916	0·856
r_{AB} (intraclass)	0·701	0·710	0·896	0·786

These values indicate no significant differences between the correlations pertaining to annual and biennial inflorescences. They do suggest a rather high correlation between the number of bracts and rays on first and second branch primary umbels.

The relation between floral parts on B and T and A is manifested by the following multiple and partial correlation coefficients.

Description	Bracts		Rays	
	Annuals	Biennials	Annuals	Biennials
$r_{B \cdot AT}$	0·807	0·796	0·683	0·871
$r_{BA \cdot T}$	0·636	0·656	0·662	0·680

Again there are no significant differences between these coefficients indicating that the amount of relationship remains the same between these floral parts pertaining to annual and biennial inflorescences.

Suggestions for further study

It might be argued that biennial plants should naturally be larger in every way since these plants had a longer time in which to establish themselves than the annuals; that the root system of the second season plants are much better for supporting the plants than those of the first season. This may be true. To overcome this criticism and to make more reliable comparisons between annual and biennial plants and inflorescences it might prove of real value to secure seeds from one plant as done in this study, save seeds from the annual and biennial flowers, and plant these seeds under the same environmental conditions and then make comparisons between the counts made in this study. Seeds should be planted in the fall and in the spring. Investigations along these lines may produce more interesting results

Summary

This study has shown that:

1. The average number of branches on biennial inflorescences of *Daucus carota* is larger than the average number on annual inflorescences.

2. The average number of inflorescences on biennials is larger than on annuals.

3. The average number of bracts per biennial clusters is larger than the average on annual clusters.

4. The average number of primary rays on biennial umbels is larger than that on annual umbels.

5. The correlation coefficient between bracts and rays is about 0·60 for annual and biennial clusters.

6. The size of plants and seasons (first and second) do not affect the amount of correlation between floral parts on stem terminals and branch terminals.

7. The amount of correlation between certain floral parts on branch terminals and non-terminals is about the same for annuals as biennials.

8. The amount of correlation between bracts and rays pertaining to first and second branch primary umbels is about the same for annuals as biennials.

REFERENCES

Baten, W. D. (1933). "A statistical study of *Daucus carota* L." *Biometrika*, **25**, 186–95.

—— (1934). "A statistical study of *Daucus carota* L. II." *Biometrika*, **26**, 443–68.

BOOKS RECEIVED

Modern Machine Calculation with the Facit Machine Model Lx. By H. SABIELNY; translated and revised by L. J. COMRIE and H. O. HARTLEY. London: Scientific Computing Service, Ltd., 23 Bedford Square, W.C. 1. 1939. Price 5*s*.

Statistical Method from the viewpoint of Quality Control. By WALTER A. SHEWHART; edited by W. EDWARDS DEMING. Washington: The Graduate School, Department of Agriculture. 1939.

Theory of Probability. By HAROLD JEFFREYS, F.R.S. Oxford: At the Clarendon Press. 1939. Price 21*s*.

The Probability Integral. By the late W. F. SHEPPARD, F.R.S., being Vol. VII of the British Association's series of Mathematical Tables. Completed and edited by the Committee for the Calculation of Mathematical Tables. Cambridge: At the University Press. 1939. Price 8*s*. 6*d*.

The Races of Europe. By CARLETON STEVENS COOK. New York: The Macmillan Company. 1939. Price 31*s*. 6*d*.

Tuberculosis and Social Conditions in England, with special reference to Young Adults. (A statistical study.) By P. D'ARCY HART and G. PAYLING WRIGHT. London: National Association for the Prevention of Tuberculosis. 1939. Price 3*s*.

Statistical Procedures and their Mathematical Bases. By CHARLES C. PETERS and WALTER R. VAN VOORHIS. New York and London: McGraw-Hill Book Company, Inc. 1940. Price $4.50.

Vol. XXXII. Part II October, 1941

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

Reprinted by offset-litho, 1952, 1960

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[*Issued* 31 *October* 1941]

THE LAWS OF CHANCE, IN RELATION TO THOUGHT AND CONDUCT

INTRODUCTORY, DEFINITIONS AND FUNDAMENTAL CONCEPTIONS

BEING THE FIRST OF A SERIES OF LECTURES DELIVERED BY KARL PEARSON AT GRESHAM COLLEGE IN 1892

[It is just fifty years since Karl Pearson took up the part-time appointment of Lecturer in Geometry at Gresham College in the City of London. This appointment, which he held during the years 1891–4, involved the delivery of certain courses of public evening lectures. His first course on 'The scope and concepts of modern science', commenced on 3 March 1891; much of its material was afterwards published as *The Grammar of Science*. Later series of lectures dealt with 'The geometry of statistics' and 'The laws of chance'. The lecture printed below was found among Pearson's papers; it was delivered on 1 November 1892 and was the first of a series devoted to the theory of probability.—ED.]

IN everyday life we feel, and justifiably feel, irritated with the man who is perpetually asking us to define the words we use. We are wont to reply that we use our terms in the 'ordinary' or 'customary' sense. As a general rule mankind understand each other in ordinary intercourse and do not stop to discuss the meaning of words. But in important and delicate business or in legal contracts the accurate definition of the words employed becomes of the utmost weight. Even more urgent still is clear definition in the matter of scientific investigation. It will not do here to appeal to that vague or floating sense of a word, which is termed the 'ordinary' or 'customary' one, for hardly any two persons use the same abstract word for precisely the same range of ideas. What is still more remarkable is the change which the meaning of words undergo in a few generations, so that even the language of our grandfathers requires to be read in the light of their (and *not our*) customary use of words. Take words apparently so simple as Nature, Right, Belief, Law, Chance: what a gulf separates the field of ideas we associate with these terms in 1892, from that which was their 'ordinary' or 'customary' value some century ago, i.e. in the days of the French Revolution! Or, again, how different is the modern scientific use of the words 'nature' and 'law' from the sense often to-day put upon them in popular or current language! Indeed I am inclined to think that the irritating person, who insists in everyday life on definitions, is after all rather a social blessing than a social nuisance—for in my experience 90% of the wordy discussions which arise in ordinary life are due to the fact that the disputants have not first fixed the sense in which they are using some fundamental conception.

In the present course of lectures, which will deal with the theory of probability, with chance, luck and the vexed question of the scientific measurement of belief, we shall have to be especially careful that we clearly define and appreciate our

fundamental conceptions. This insistence on definition must be the starting-point of any really scientific discussion, and I want to urge you all to start the study of this or any other subject by trying to clearly define its scope and terms. You must in this respect 'list to what the friar preaches and not to what he does', for according to a sharp-eyed reviewer I have myself been guilty of publishing a book, in which no definition of *chance* itself was given! I will endeavour to supply that omission in to-day's lecture. But first I want to point out the relation between the subject of my present course and the topics of the two earlier ones on the *Fundamental Concepts of Science* and on *Statistics.* The relationship is a very close one indeed, far closer than might be imagined on a cursory examination. We shall find that statistics are the practical basis of much, if not all scientific knowledge, while the theory of chance is not only based on past statistics but enables us to calculate the future from the past, the very essence of scientific *knowledge*. There is a close relation between *provable* and *probable*; the analysis of 20,000 tosses of a coin will help us to penetrate into the very laboratory of Nature whose complexity presents us with results strikingly akin to those of a game of chance; while the record of a month's roulette playing at Monte Carlo can afford us material for discussing the foundations of knowledge. That things apparently so diverse should be so closely related may strike some of you as paradoxical, and indeed the ground we are to venture upon abounds in difficulties and pitfalls. It is one where criticism and controversy have been very rife, but at the same time have been fruitful of results and have contributed to clearness of thought. We shall find well-marked divergencies of opinion, characterizing two different schools, which push to extremes in different directions. Based on the researches of Laplace and Quetelet, we find De Morgan, John Stuart Mill and Stanley Jevons pushing the possibilities of the theory of probability in too wide and unguarded a manner; while in the opposite camp we find George Boole and Dr Venn taking a severely critical and in some respects perhaps too narrow view of them. As in many other cases the safe road is probably the middle road, and this road is that which I conceive Prof. Edgeworth of Oxford to have pointed out. For those of you who may have time for reading I would strongly recommend a comparison of Chaps. X–XII of Stanley Jevons' *Principles of Science* with Chaps. VI–XI of Dr Venn's *Logic of Chance* and Prof. Edgeworth's *Philosophy of Chance* published in *Mind* for 1884. I shall refer to the opinions of these writers in the course of our work, but you would find the subject of chance as treated by them enticing in the extreme, and they will give you far more amply than I can do in these lectures the various features of the controversy.

While dealing with the subject of books I may also refer to:

DE MORGAN: *Formal Logic* (1857). Here Chaps. IX–XI are closely connected with the topics of our first two lectures.

DE MORGAN: *An Essay on Probabilities* (1838). This is still a useful and suggestive little book, although it requires some mathematical knowledge.

WHITWORTH: *Choice and Chance* (3rd ed. 1878). An excellent book with which to approach the elements of the mathematical theory.

WESTERGAARD: *Die Grundzüge der Theorie der Statistik* (1890). By far the best textbook on the relation of *statistics* and probability for those *who read German*.

Now I want to restate in the first place some of the conclusions I placed before you in my first course of Gresham Lectures. I do not want you now—any more than I did then—to accept those conclusions as your own but rather to probe and investigate them for yourselves, and thus ascertain whether they form a basis sufficiently sound for the superstructure placed upon them. The conclusions to which I want to draw your attention are those concerning the material of science, scientific law and cause and effect. In the first place the material of science consists of certain groups of sense-impressions, which we term phenomena and in which we mark not only a certain permanency but a routine. When we find a certain sequence of sense-impressions frequently repeating itself, we speak of any antecedent sense-impression as a cause, any subsequent one as an effect. A, B, C, D, E, F being a succession of sense-impressions, which repeats itself, A, B, C, D, E are all termed *causes* of the *effect* F. A scientific law or formula is a statement which enables us to resume or describe in brief language a routine sequence—or many such routine sequences—of causes and effects. As I pointed out to you, a scientific law does not *enforce* a sequence, it merely describes what takes place. No law of gases *causes* or *enforces* the boiling of a kettle of water, when placed on the fire; it merely describes *how* it boils. How then do we know that a kettle of water will boil, if placed on the fire? The answer is a very simple one, our knowledge of what will happen is based upon *past experience*. *Statistics* of past experience, our own or that of other men, are the basis of our knowledge of all cause and effect, of all knowledge of phenomena. Here you have the kernel to statistics as the basis of knowledge. The statistics are formed in a rough practical manner, but are none the less real for all that. Take any sequence of phenomena such as a kettle of water boiling which has been long enough on the fire. We have behind us the invariable experience that kettles in like positions do boil, and we say we *know* that this kettle will boil. If it does not we expect that some portion of the customary sequence fails, the fire has gone out, there is no water in the kettle, or there is somewhere a breach in the 'group of causes'. But in our statement about the kettle there are really two important factors, there are the statistics of past experience, and the assumption that these statistics will apply to the future. There is no scientific reason why the same groups of causes should always be followed by the same effect. Indeed, a distinguished American mathematician, Mr C. Pierce, has gone so far as to support the view that the causes, A, B, C, D, E, may be followed by F or G, indifferently. There is no logical or intellectual proof that like causes will be followed by like effects. It is purely a

result of experience. Statistics show us the prevalency of routine in the past, and these statistics are the first basis of our knowledge. That what has held in the past, will hold for the future, is again a statement for which there is no proof; it is the outcome of our experience of what has happened in the pasts, which were at an earlier date futures. Hence our inferences with regard to natural phenomena are essentially based on statistics—namely statistics of what has happened in the past, and the experience that within certain ranges the statistics of the past repeat themselves in the future. Now I want you to grasp this point very clearly, for we are coming close to the relationship between statistics, knowledge, belief and chance. What do we mean when we say that 106 boys are born as compared with 100 girls? Or when we assert that such will take place next year? Why simply this that the statistics of past years for a very great number of births give us boys and girls repeatedly in these proportions, and further experience—in other words statistics again—shows us that such statistical ratios do not change suddenly and abruptly, the results calculated for a period of four or five years, hold very closely for the following four or five years. Or, again, when we say that we *know* that the sun will rise to-morrow—we are just as much appealing to past experience of the action of the sun and past experience of the occurrence of routine, as when we appeal to the statistical appearance of births. The law of gravitation does not enforce the rising of the sun, it is merely a scientific description of what we observe in the motion of the planets. Suppose the sun had not risen on one well-authenticated occasion in our experience, and on one only, we should then be slightly less confident in our assertions as to its appearance to-morrow. Our knowledge would then have been weakened down into some very strong form of *belief*. The more frequently the sun had omitted to rise, the less strong would be our certainty with regard to its conduct to-morrow, until we passed through every shade of belief to disbelief itself. Or let us take a more tangible and possible case. A friend is leaving us, say in Chancery Lane at 4 o'clock in the afternoon, and we tell him that he will find a Hansom cab at the Fleet Street corner. There is no hesitation in our assertion. We speak with knowledge, because an invariable experience has shown us Hansom cabs at 4 o'clock in Fleet Street. But given the like conditions within reach of a suburban cab-stand, and our statement becomes less definite. We hesitate to say absolutely that there will be a cab: 'You are sure to find a cab', 'I believe there will be a cab on the stand', 'There is likely to be a cab on the stand', 'There will possibly be a cab on the stand', 'There might *perhaps* be a cab', 'I don't expect there'll be a cab', 'Its very improbable', 'You are sure not to find a cab', etc., etc. In each and every case we go through some rough kind of statistics, *once* we remember to have seen the stand without a cab; on occasions few and far between, 'perhaps on an average once a month', 'perhaps once a week', 'every other day', 'more often than not there has been no cab there'. Certainty in the case of Fleet Street passes through every phase of belief to dis-

belief in the case of the suburban cab-stand. If once a month is the very maximum of times I have seen an empty cab-stand, my belief that my friend will find a cab there to-day is far stronger than if I have seen it vacant once a week. A measure of my belief in the occurrence of some event in the future is thus based upon my statistical experience of its occurrence or failure in the past. When in a wide range of experience there has been no experience of failure, then as in the case of the cab in Fleet Street, or in the case of the sun rising to-morrow our belief becomes so strong that we speak of *knowing*. But all this knowing really amounts to is a very high, or even the highest possible, degree of *probability*. I *know* that the three angles of a triangle together make two right angles, for this lies in my definition of triangle, and belongs to the field of mental conceptions and not to physical phenomena. But of the physical universe I can only say I *believe* such and such things will occur, and the degree of my belief is measured *in a rough approximate way* by the statistics of past occurrence and failure.

We can see this better, I think, by returning to the definite case of the cab on the stand. Once a week on the average of a long experience I have seen the stand empty. Thus for every six occasions there is a cab, there is one occasion that there is not a cab. Had I sought for a cab at the given hour for a long period I should have been successful six times in every seven. We then define the ratio of the number of *successful* instances to the total number of occasions as the *probability* or *chance* of finding a cab—in this case the chance is 6/7. Thus the chance of an event is the numerical measure of past experience. It is based essentially on *statistical information*. How wide must be the range of information on which the chance is based we will consider later, for it involves very many important points. At present we have the following simple rule: Taking the statistics of the occurrence, find the number of favourable instances and divide them by the total number of instances and this is the chance of the event. Returning to our cab-stand, suppose that only once in four weeks I have seen it empty at 4 o'clock on the average, then the chance of the event, finding a cab at 4 o'clock is 27/28—i.e. in the long run 27 favourable instances per 28 occurrences.

Now, if I have found only one failure in 28 occurrences my hope of finding a cab on a particular occasion—my *belief* in there being a cab—will be far greater than if I have found a failure once in 7 occasions. Thus my belief is in some way related to the chance; if I know the chance is greater, my belief is greater. Prof. De Morgan has asserted that the proper measure of belief *is* chance, and according to him my belief in the two cases cited above would be as 6/7 to 27/28, or as 8 to 9. He thus reaches an exact numerical appreciation of belief, what might be termed a scientific measurement of belief.

This view of De Morgan's has been severely criticized by Dr Venn. He asserts that chance is something objective or physical—is based on statistics of the occurrence of a physical phenomena—while belief is something psychical and is

largely determined by the emotional and nervous temperament of the individual man. In other words it is subjective and not objective. According to Dr Venn probability deals with the laws of things, while according to De Morgan probability has to do with the laws of our thought about things.

Now I think we must agree with Dr Venn that it is impossible to set an absolute numerical value upon the beliefs of human beings in practical life. No one will venture to say that one of his beliefs is exactly nine times as strong as another. Perhaps the only practical measure we can form of the strength of beliefs is the readiness of men to act upon them, and the impetuous or credulous man will risk as much where the chance is small, as the sluggish or sagacious man where the chance is large. At an important crisis he will risk finding a cab on a chance which would have induced the prudent man to order one beforehand. Clearly then as applied to the beliefs of practical men in actual life, Dr Venn is right in asserting against De Morgan that we cannot put an exact numerical value on belief. On the other hand I think we must question whether chance can conveniently be treated as peculiar to things. The means by which statistics are taken in practical life are human and they become subjective and individual in the process of taking and applying them. Besides this the statistics on which the chance may be reckoned are frequently at the option of the particular individual and the chance at once becomes subjective and peculiar to him. Let me point out what I mean. A man is tossing a coin in a railway carriage, a country lad in the carriage, *judging by his experience of coins in the past*, is ready to believe that the chance of a head is 1/2, i.e. that once in two occasions in the long run it will come down head. A scientific man (also without guile!) who has made experiments in tossing coins knows that every coin has a slight bias, and that there is in all probability a slight fraction of a per cent more heads or tails in the long run in the tossing of this particular coin. A man of the world knows that the coin-tosser is a swindler, and judges that his coin is loaded, so that the chance that it comes down head is very far from a half. And the coin-tosser himself? Well, he has no statistics at all of the conduct of this particular coin—it may be true or biased or loaded—but being an adept in tossing he can bring it down head or tail as he pleases. What are we to say is the *chance* that this coin will come down head? We have no statistics whatever of what happens when swindlers toss coins, which may after all unknown to them be loaded! Are we to say that the chance is an unknowable quantity, and that we cannot make any application of the theory of probability? I am inclined to think this would unduly narrow the field of our science. It seems to me that we can and should apply our theory to the chances subjectively estimated of each occupant of the railway carriage. These chances are based on the subjective experience of each individual with regard to coins under like conditions, and they certainly are more concerned with the laws under which people think about things, than with the laws of things themselves. If the country lad bets on a head the chance

of a head is for him one-half, for the scientific gentleman it must be a shade less or a shade more, for the man of the world it is a very small chance indeed, for the swindler it may be a certainty, if he means the lad to win on the first occasion, in order to excite him to betting heavier amounts. Now the beliefs of these four persons clearly differ in strength and their relative proportions are closely related to the individual measurement of the chance, to what we may term the *subjective chance*. I am inclined to think with De Morgan that belief varies very closely with the subjective chance, *but*, this subjective chance depends upon the statistics of individual experience, and may differ widely from what we may term the *objective chance*, or the chance based upon statistics of the actual event in question and independent of the individual calculator.

Turn (I hope, for the last time) to our cab-stand and the chance of finding a cab on it. Accurate statistics may have been taken of the absence of cabs upon it for a long period, perhaps, two or three years. For our present purposes these may represent the statistics for the calculation of the 'objective chance'. I may have observed the cab-stand, not *very* regularly, for a few months, and my result is: empty about once a week at 4 o'clock; a friend knows nothing about this particular cab-stand, but has formed statistics of suburban cab-stands in general; while another person without paying special attention to suburban cabs has formed pretty precise ideas as to cab-stands in London as a whole. The statistics of suburban cab-stands in particular, or of London cab-stands in general may be wide and accurate, or may be individual and approximate; in either case it is a *subjective act* which classes the particular cab-stand under either of these headings, the particular chance selected is the result of individual experience or *subjective choice*. If we ask what is the relation between subjective chance and objective chance, I think we can safely say, that while the two often differ widely, yet the more deep a man's experience, the more thorough his observation and his knowledge of phenomena, the more closely his subjective statistics will fit the objective statistics. He will never, perhaps, make the two coincide, but in the long run of practical life his mistakes will be few and tend to balance each other; his subjective chance will approximate to the objective chance in Dr Venn's sense. He will know what classes of statistics to apply to individual cases with the best results, or in ordinary language, 'he will be a judge of men and things'. If experience of life and acquaintance with fact lead a man's subjective appreciation of chance to approximate to the objective value of chance, may we not say that, if belief varies with a man's subjective view of chance, then ultimately it is objective chance which governs belief?

There is a difficulty here which is I think sometimes overlooked, but which seems to me fundamental and we must regard it with a little care. I take a coin and I say the chance, when it is tossed, of a head is one-half. Now what exactly does this mean? One or other of two things, either:

(1) I have tossed this same coin 10 or 20,000 times and found practically the

same number of heads and tails. Here the subjective and objective chances are practically identical. Or:

(2) I have not tossed this special coin at all, but judge it to be like other coins, of which my own rough experience, and that of other men, presents practical statistics of the equality in the number of heads and tails obtained in a great number of tosses.

Here the subjective and objective chances may or may not be the same, for after all the coin may be loaded or even a double-headed one. But in this case also experience of the coin will ultimately bring the subjective and objective chances to the same value, be it 1/2 or otherwise. Now let us go a stage further and suppose that experience has brought the subjective appreciation of chance to its objective value. Would that objective value be a measure of my belief? Now there is an assumption here, which I have before referred to, and want you now to particularly notice. The chance is really based on past statistics, it is the number of successes observed by the total number of trials. 20,000 times the coin has been tossed and 10,000 times—within a few units, perhaps—heads have appeared; the chance of head is $\frac{10,000}{20,000}$ or 1/2. But this is not all we mean when we say the chance is a half. We refer to the future as well as to the past, and we *assume* that if we were to throw the coin an indefinite number of further times, there would be in the long run as many heads as tails. Here is the assumption we make when *chance* is taken *as the basis of belief as to the future*. In other words the statistics of past experience are assumed to be identical with the statistics of what will happen in the future. When I say that the chance of a head is one-half, that statement is meaningless, if it be considered as referring to a single toss, it refers to what *I believe* will happen on the average in a very great number of future throws—i.e. a practical equality of heads and tails. This belief is based on two elements, first, statistics of past experience as to tossing and secondly the permanence of statistical ratios.

This latter is a most important element and one which in reality forms a large factor of belief. Let us bring this out more clearly by a comparison of one or two cases. Statistics of tosses show a coin to be a true coin, to give in the long run head as often as tail—chance of a head therefore 1/2.

Statistics of a certain country show that of 206 births 106 are boys—chance of a boy being born 106/206. My statistical experience of a certain cab-stand, shows that on an average there is no cab there once a week at 4 o'clock—chance of a cab 6/7.

Now before I apportion my belief of what will happen in the long run in the future in these several cases, I have to consider the *permanence* of these statistical numbers, and the only way I can do this is by examining statistics as to the permanence of similar numbers. The factor of my belief depending on the permanency of the chances is itself rooted in statistics. How often have I found

the chance given by statistics to be constant, how often to change? What indeed is the 'chance' of a chance changing?

A coin has given as many heads as tails in the past, why should it not now begin to give a vastly greater proportion of heads? The appeal is again to experience, and experience tells us that, if a coin be not battered, bent or altered, it maintains indefinitely the same chance of a head.

On the other hand experience tells us that while in vital statistics there is scarcely ever an abrupt change, ratios do alter slowly and gradually, the chance of a boy being born as calculated from the last few years may hold for the next few, but it may vary from decade to decade and century to century.

Still more may the chance of finding a cab on the stand vary. I may have carried out my observations for two or three months, but the completion of a new line of railway or a Licensing Act may make a sudden breach of continuity, there is much less permanence in statistics of this kind, than in those of coins or babies. Clearly the chance determined from past statistics is *not* the only factor in apportioning my convictions as to the future appearance of heads, boy babies, and cabs. The chance of the statistics remaining in the future what they have been in the past must also be considered and be shown to be the same in all the cases where beliefs are compared.

Thus, I think, we must agree with Dr Venn, although partly on other grounds, in recognizing that the chance of an event is not an accurate numerical measure of our belief in its occurrence, but on the other hand we may go so far with De Morgan as to assert that our belief is strengthened or weakened when the subjective chance based on our personal knowledge or experience—or on the rough and ready statistics of practical life—is increased or decreased.

We may even go a stage further and construct a model universe in the following manner: Suppose a world in which men had such width of experience that their subjective appreciation of chance was equal to its objective value, and further that in this ideal world statistical ratios retained a permanent value—in the manner in which we actually find they do in games of chance—then in such a scientific ideal world chance might fairly be considered to measure belief.

It may be asked what is the use of such an ideal model as this? In the flesh and blood men of actual life with their prejudices and half-knowledges the subjective appreciation of chance diverges often widely from its objective value; further in this real world few statistical ratios are actually permanent, they vary not only with time, but with the range and limits of our statistics. What then can the use of our model be? Well, of much the same use as the political economists' model of society governed by the laws of exchange or value, or the physicists' molecular model of nature. Neither is true to reality, but both serve with certain reservations to describe in broad terms the general facts of economic and physical phenomena. In the same manner, because in a rough and approximate way men's subjective appreciation of chance does tend in the practical

experience of life to approach the objective value—and because in a great variety of cases chances calculated on past experience are found to remain permanent in the immediate future—so men's beliefs as evidenced especially in conduct do vary with chance; and if chance be not a scientific measure of belief, it is yet in the average of men a rough and ready means of gauging, more or less accurately, the relative strength of convictions.

It may seem strange to some of you to be told that chance is the measure of past experience. At first there may appear to be a very considerable difference between the chance that a boy or a girl will be born and the chance that a head or tail will turn up. We are quite ready to admit that statistics are needful in order to determine whether more boys or girls are born and so to determine the chance of a boy or girl birth. But we are not inclined at first to admit an equal necessity in the case of the coin. We are inclined to argue that 'We *see* no reason why head should occur more frequently than tail' and then convert this into 'There can be no reason why head should occur more frequently than tail'—and then 'Head and tail must be equally frequent'. You will see the weakness of this argument at once by applying it to the case of boys and girls. 'We see no reason why more boys should be born than girls.' 'There can therefore be no reason why more boys should be born than girls,' and finally 'No more boys are born than girls'. Here statistics step in and upset all our preconceived notions. In fact all arguments of this kind remind us of the old mediaeval notions of physical science, which began by arguing as to what nature *ought* to do, instead of patiently observing what she *did* do. We may *see* no reason why head should occur rather than tail. But if there were not a very definite reason why head or tail should have the preference in each individual throw, then we may be quite sure that the coin would balance on its edge and exhibit neither head nor tail.

Mere inspection of a coin would certainly not suffice to tell us that the 'chances' of head and tail are equal. The head is different in shape and appearance from the tail, and the coin may really be biased by this. Let us get over this difficulty by taking a perfectly uniform disk absolutely alike on both sides, and let us imagine it thrown up so that neither side has any advantage either on leaving our hand or on reaching the ground. Can we say that the chances of either side are equal without any appeal to experience—to statistics? The fact is that if the two events were absolutely balanced in this manner, if not only we *saw* no reason why one should occur more than the other, but there *was* no reason, then there would be no *chance* of either event occurring at all. In our experience of nature there is no such thing as chance of this kind. The moment a coin or a die leaves the hand, its fate is really settled and there is no field for the 'play of chance' in the obscure sense we have just been referring to. The mechanical causes are perfectly definite and the occurrence of head or tail, ace or deuce, absolutely certain. It is quite true that these mechanical causes

are far too complex, too evenly balanced and too incapable of measurement for us to mechanically describe what must happen, and so predict head or tail. Mechanically the one or other is predetermined, but the multiplicity of causes varying so slightly and yet so effectively from throw to throw leaves us in ignorance as to the result. If we are merely ignorant as to a result which is mechanically perfectly certain, what is the meaning of chance in physical nature? Simply this that we aid our ignorance by an appeal to past statistical experience. The chances of a coin falling head or tail being equal does *not* depend on my ignorance of what will occur, or on my *seeing* no reason why head more than tail should occur, but on my *experience of the statistics of tossing coins*. This experience is really summed up in the symbolic slang 'a toss up'—as an expression for an equality of chances. I know from my own personal experience and from the common habits of men—as well as from the statements of gamblers and others—that loaded coins are not of frequent occurrence. Without this experience I could predict nothing of the tossing of a coin, it might invariably come down 99 % head; or having fallen on the first occasion head or tail, that fall might in itself determine what it would do on the second occasion.

What I have said of tossing a coin holds good for the drawing of black and white balls out of a bag. It might seem at first sight that if 50 white and 50 black balls were put into a bag and well mixed, then, each ball being replaced after the drawing, as many white as black balls will be drawn in a large number of trials. In other words the chances of drawing white and black balls are equal. But here again, if our statement is really to mean anything we must be appealing to some rough experience of the conduct of balls in bags. It is conceivable that the hand might have a preference for black balls, or that white balls would have a preference for each other and the bottom of the bag. If it be objected that the hand does not detect colour difference, and that gravity acts equally on equal balls if they be of different colours, we are at once appealing to a wide statistical experience resumed in certain fundamental laws of nature. We have left at once the shaky ground of subjective reasoning, and turned to statistics.

But even in these cases direct experiment comes to our aid and provides the statistics, which are only roughly embodied in the everyday experience and opinions of mankind. Thus:

BUFFON tossed a coin 4040 times, there resulted: 1992 heads, 2048 tails, or 49 % heads, 51 % tails.

QUETELET made 4096 drawings out of a bag containing an equal number of black and white balls, there resulted: white balls 2066, black balls 2030, or 50·4 % white and 49·6 % black.

MR GRIFFITH, one of my students, has kindly tossed a penny 8178 times and there resulted: 4092 heads, 4086 tails, or 50·04% and 49·96 % tails.

WESTERGAARD made 10,000 drawings out of a bag containing equal numbers of red and white balls well shaken before each drawing after the replacement of

the previously drawn ball. He obtained: white balls 5011, red balls 4989, or 50·11 % white and 49·89 % red.

YOUR PRESENT LECTURER tossed (as a holiday task) a shilling 24,000 times. He obtained for the first 12,000 tosses: 5981 heads, 6019 tails, or 49·84 % heads, 50·16 % tails; second 12,000 tosses: 5992 heads, 6008 tails, or 49·933 % heads, 50·067 % tails. In both series there is a balance in favour of tails: in the first of less than 1/6 %; in the second of about 1/15 %.

Taking both series together we have: 24,000 tosses: 11,973 heads, 12,027 tails, or 49·8875 % heads and 50·1125 % tails.

To avoid any chance of there being a slight loading in the coin—a very slight bias towards tails—let us call the heads, tails and tails, heads in the first 12,000 tosses, we then find: 12,011 heads and 11,989 tails; or 50·046 % heads and 49·954 % tails.

Thus to 1/20 of a per cent heads and tails are equal, or to express it in another manner there has on the average been only one head too many in 1200 tosses. Buffon's experiments coincide with mine in showing a slight bias in favour of tail.

Finally I have analysed the red and black events in an entire month's play of the roulette tables at Monte Carlo. I find that out of 16,178 throws of the ball* 8111 fell into a red number and 8067 into a black, or there were 50·14 % red and 49·86 % black.

These experiments amply confirm the rougher statistical experience of mankind as to the equality of chances in tossing, or drawing balls from bags, or playing roulette. *It is on experience of this kind, on accurate statistical measurement, not on a priori reasoning or subjective opinion, that the data of probability are to be based.*

In my next lecture I shall deal more at length with the nature of the statistics by which we supplement our ignorance of what is about to happen.

* The twenty-seventh figure, 0, was of course omitted to equalize chances.

MEDICAL STATISTICS FROM GRAUNT TO FARR

By MAJOR GREENWOOD

INTRODUCTION

UNDER the Fitzpatrick Trust, a Fellow of the Royal College of Physicians of London is chosen annually by the President and Censors to deliver two lectures in the College on 'The History of Medicine'. I had the honour of being chosen for this office in 1940 but, for obvious reasons, the lectures were not delivered, and it may be safely assumed that some years will pass before a medical audience will have time to attend to the history of a subject the modern practice of which does not make a strong appeal to physicians.

The nature of the intended audience inclined me to stress the medical rather than the purely statistical aspects of the story and I have trodden ground over which a greater man passed some years ago. I hope that Karl Pearson's studies of some or all of these old heroes will eventually be printed, and I know that my slight essays can ill sustain a comparison. But, precisely because they are slight and linger over small traits and human oddities, they may, in these times, wile away an hour or two. I have eliminated some explanations which no statistician or biometrician needs and the medical technicalities are few. Perhaps a note on the London College of Physicians as it was in the days to which these studies relate should be added.

The College was more than a century old when John Graunt was born, and the corporation consisted wholly of physicians who were Doctors of Medicine of Oxford or Cambridge; these were the *Fellows*. Physicians not Doctors of Medicine of Oxford or Cambridge were admissible only to the grade of *Licentiate*, and it was not until the nineteenth century, when Farr was a young man, that the exclusive privilege of the senior universities was abolished. It was not until Farr was a middle-aged man that the College had any direct contact with general practitioners of medicine and began to examine persons who did not seek to practise solely as physicians. In modern usage the College licence, L.R.C.P. (now only granted jointly with the membership of the Royal College of Surgeons, M.R.C.S.), is a diploma obtained by a large proportion of general medical practitioners in the South of England. Down to Farr's time, the L.R.C.P. was a 'specialist' diploma and could not have been taken by a general practitioner (the apothecary of those days) at all. The old L.R.C.P. is represented by the M.R.C.P. of our own time but with this distinction. Now, Fellows (F.R.C.P.) are normally chosen from the body of M.R.C.P.'s. In the past only Doctors of Medicine of Oxford or Cambridge could be Fellows, and before election but after examination were known as 'candidates', not licentiates. The great physician

Sydenham was never more than a licentiate. He graduated M.B. at Oxford and, for some unknown reason, never proceeded M.D. until near the end of his life, when he took the higher degree not at Oxford but at Cambridge.

I. THE LIVES OF PETTY AND GRAUNT

It is always rash to assign an absolute beginning to any form of intellectual effort, to say that this or that man was the very first to fashion some organon which has proved valuable. All we are justified in saying is that this or that man's work can be shown to have so directly influenced the thought of his contemporaries or successors that from his day the method he used has never been forgotten. It may be that the lost works of the school of the Empirics Galen despised anticipated the numerical method of Louis—some words of Celsus are consistent with the hypothesis. It may be that in the long succession of parish clerks who for more than a century transcribed the London Bills of Mortality, one or two suggested that these figures might have some other use than that of warning His Highness of the need to move into Clean Air. But we do not know. We do know that out of the casual intercourse of two Englishmen in the seventeenth century was produced a method of scientific investigation which has never ceased to be applied and has influenced for good or ill the thought of all mankind. In that sense at least we may fairly hold that John Graunt and William Petty were the pioneers not only of medical statistics and vital statistics but of the numerical method as applied to the phenomena of human society.

John Graunt and William Petty were both of Hampshire stock. Petty was of Hampshire birth, born on Monday, 26 May 1623, and was three years younger than John Graunt, who was born at the Seven Stars in Birchin Lane on 24 April 1620.

Materials for writing Petty's life are abundant; indeed a good biography of him was written nearly fifty years ago by his descendant Lord Edmond Fitzmaurice, and since then much of the material used by Lord Edmond has been printed. Sources for Graunt's biography are scanty, the most valuable John Aubrey's brief life of him.* Graunt and Petty became acquainted in or before 1650. The circumstances of that first acquaintance are interesting to those who meditate upon the perepeteia of human fate. It was the contact of client and patron.

John Graunt's early life and manhood were those of the Industrious Apprentice. His father was a city tradesman, who bred his son to the profession of haberdasher of small wares. John 'rose early in the morning to his study before shop-time', and learned Latin and French, but did not neglect his business. He was free of the Drapers' Company and went through the city offices as far as

* *Brief Lives, chiefly of Contemporaries*, set down by John Aubrey, between the years 1669 and 1696, edited by Andrew Clark, Oxford, 1896, 1, 271 *et seq.*

common councilman; he was captain and then major of the trained bands (the ancestor of the Honourable Artillery Company). At the time of the Great Fire he is said to have been an opulent merchant. Even fifteen years earlier he—and no doubt his father (1592–1662)—had city influence. At that time a Gresham professorship was vacant and a young Dr Petty was anxious to obtain it. This young man's career had been unlike that of an industrious apprentice; it had been, even for the seventeenth century, romantic. His father was a clothier in Romsey, who 'did dye his owne cloathes' in a small way of business. When William was a child, 'his greatest delight was to be looking on the artificers—e.g. smyths, the watch-maker, carpenters, joyners etc.—and at twelve years old could have worked at any of these trades. Here he went to schoole, and learnt by 12 yeares a competent smattering of Latin, and was entred into Greek' (Aubrey, Clark's edition, **2**, 140).

But the precocious lad did not find a patron in Romsey and was shipped for a cabin boy at the age of fourteen. His short sight earned him a taste of the rope's end, and after rather less than a year at sea he broke his leg and was set ashore in Caen to shift for himself. 'Le petit matelot anglois qui parle latin et grec' attracted sympathy and obtained instruction in Caen. Caen was not a famous seat of learning like Leyden or Montpellier, but the Fellows and licentiates of the College of Physicians admitted between 1640 and 1700 include the names of four persons who studied or graduated in Caen (Nicholas Lamy, Theophilus Garencières, John Peachi and Richard Griffiths). Petty, however, was not then thinking of medicine but mathematics and navigation and came home to join the navy. In what capacity he served is unknown; he merely says (in his Will) that his knowledge of arithmetic, geometry, astronomy conducing to navigation, etc., and his having been at the University of Caen, 'preferred me to the King's Navy where at the age of 20 years, I had gotten up about three score pounds, with as much mathematics as any of my age was known to have had'. His naval career was short, for in 1643 he was again on the continent. Here he wandered in the Netherlands and France and studied medicine or at least anatomy. He frequented the company of more eminent refugees, such as Pell and Hobbes, as well as that of the French mathematician Mersen. He was very poor and told Aubrey that he once lived for a week on three pennyworth of walnuts, but on his return to England the three score pounds had increased to seventy and he had also educated his brother Anthony.

At first Petty seems to have tried to make a living out of his father's business, but he soon went to London with a patented manifold letter writer and sundry other schemes of an educational character. These occupied him between 1643 and 1649 and made him acquainted with various men of science, among others Wallis and Wilkins, but were not remunerative, and in 1649 he migrated to Oxford.

Petty was created Doctor of Medicine on 7 March 1649 by virtue of a

dispensation from the delegates (no doubt the parliamentary equivalent of the Royal Mandate of later and earlier times). He was also made a Fellow of Brasenose and had already been appointed deputy to the Professor of Anatomy. He was admitted a candidate of the College of Physicians in June 1650 (he was not elected a Fellow until 1655 and was admitted on 25 June 1658). At Oxford he became something of a popular hero by resuscitating (on 14 December 1651) an inefficiently hanged criminal, who, condemned for the murder of an illegitimate child, is said to have survived to be the mother of lawfully begotten offspring.

Academically Petty rose to be full Professor of Anatomy and Vice-Principal of Brasenose. It is at this point (as usual the precise dates are dubious) that he became a candidate for a Gresham professorship and made contact with John Graunt.

Although, as I have said, the materials for a biography of Petty are abundant, all we know of his early years comes from himself or from friends of later life who knew no more than he told them. We have no independent means of judging the extent of his culture. There is good evidence that he knew more Latin than most Fellows of the College of Physicians know now; none that he was an exact scholar (indeed we have his own word, which I am not prepared to gainsay,* to the contrary). He was certainly admitted to friendship by some men, such as Wallis and Pell, who were serious mathematicians, as by others, such as Hobbes, who were not. But whether he could fairly be called a mathematician is doubtful. Of his medical knowledge we know little. He left medical manuscripts, but these are still unpublished; of his clinical experience we know nothing.

Petty told Aubrey that 'he hath read but little, that is to say, not since 25 aetat., and is of Mr. Hobbes his mind, that had he read much, as some men have, he had not known as much as he does, nor should have made such discoveries and improvements'. But it is at least certain that he made a favourable impression upon men who had read a good deal and that the young Dr Petty of 1650 was thought a promising man. Still it *had* been an odd career and one wonders what a steady business man in the city of London thought of it.

Why the anatomy professor who had resuscitated half-hanged Ann Green should be made a professor of music is not obvious, and if the Gresham appointments were jobs, why should the job be done for Petty? The modern imaginative historian might suggest various reasons. For instance, that Petty made a

* If No. 88 of *The Petty Papers* (**2**, 36) is a typical example of Petty's Latin Prose style, there is not much to be said for it. Here is an example: 'An dulcius est humanae naturae permultos suam potestatem in unum quendam et in perpetuum transferre, id est pendis amittere quam ipso puel deindem servare, vel paulatium et in breve tempus irogare, a seipsis demo reformondam et disponendam alioquin pro ut, mutato tam rerum quam animi indies suaserit?' Some of the gibberish may be due to the editor's failure to decipher the handwriting, but no emendation could twist this into unbarbaric prose.

conquest of Graunt, perhaps had Hampshire friends who were friends of the Graunt family, perhaps talked about political arithmetic. We have no evidence at all. If the Gresham Professor of Music *had* duties, Petty did not perform them; about the time of his appointment he obtained leave of absence from Brasenose and within a year (in 1652) had left for Ireland, where he was to be very busy for some time to come and to make, or found, his material fortunes.

Macaulay (chap. III) says that at the end of the Stuart period the greatest estates in the kingdom very little exceeded twenty thousand a year.

The Duke of Ormond had twenty-two thousand a year. The Duke of Buckingham, before his extravagance had impaired his great property, had nineteen thousand six hundred a year. George Monk, Duke of Albemarle, who had been rewarded for his eminent services with immense grants of crown land, and who had been notorious both for covetousness and for parsimony, left fifteen thousand a year of real estate, and sixty thousand pounds in money, which probably yielded seven per cent. These three Dukes were supposed to be three of the very richest subjects in England.

In 1685 Petty made his Will. This Will is a curiously interesting document, because it is also an autobiography. It is rich in arithmetical statements and, like much of Petty's arithmetic, the statements may be optimistic. Petty's final casting of his accounts is in this fashion: 'Whereupon I say in gross, that my reall estate or income may be £6,500 per ann. my personall estate about £45,000, my bad and desparate debts, 30 thousand pounds, and the improvements may be £4000 per ann., in all £15,000 per ann. *ut supra*.'

The details of the calculation are perplexing enough; still if the above cited dukes *were* the richest subjects of the king and if (Macaulay) 'the average income of a temporal peer was estimated by the best informed persons, at about three thousand a year', Sir William Petty, of the year 1685, had travelled as far from the young Oxford professor of 1650 as that budding physician from the little English cabin boy who spoke Latin and Greek, in Caen, in 1638. The details of the fortune-building are not our concern. The shortest account is Petty's own in his Will. He says that by the end of his Oxford career he had a stock of four hundred pounds and received an advance of one hundred more on setting out for Ireland.

Upon the tenth of September, 1652, I landed att Waterford, in Ireland, Phisitian to the army, who had suppressed the Rebellion began in the year 1641, and to the Generall of the same, and the Head Quarters, at the rate of 20s. per diem, at which I continued, till June, 1659, gaining by my practice about £400 per annum, above the said sallary. About September, 1654, I, perceiving that the admeasurement of the lands forfeited by the forementioned Rebellion, and intended to regulate the satisfaction of the soldiers who had suppressed the same, was most insufficiently and absurdly managed, I obtained a contract, dated the 11th. of December, 1654, for making the said admeasurement, and by God's blessing so performed the same as that I gained about nine thousand pounds thereby, which with the £500 above mentioned, my sallary of 20s. per diem, the benefit of my practice, together with £600 given me for directing an after survey of the adventrs lands, and £800 more for 2 years sallary as Clerk of the Councell, raised me an estate of about thirteen thousand pounds in ready and reall money, at a time, when, without art, interest, or authority,

men bought as much lands for 10s, in reall money as in this year, 1685, yield 10s. per ann. rent above his *Maties* quitt rents (*The Life of Sir William Petty*, by Lord Edmond Fitz-maurice, London 1895, p. 319).

No one would willingly rake over the embers of Irish history—still glowing after nearly three hundred years. Petty believed himself to be a good man struggling against adversity and a public benefactor treated with gross injustice to the day of his death. Lecky (*History of Ireland*, vol. 1, chap. 1, p. 111 of popular edition) took a less favourable view. Even if the subject were relevant to my undertaking, which it is not, I have not the training in historical research to justify me in writing about it. There are, however, some points of psychological interest.

Petty did not, like his contemporary Thomas Sydenham, actually take up arms against the king, but he was even more plainly a protégé of the king's enemies. Sydenham's military career was unimportant; there is no reason to believe that he ever exchanged a word with a member of the Cromwell family. Petty was the confidential adviser and close personal friend of Henry Cromwell; his services to the Commonwealth authorities were the foundation of his fortune. Like many people who have social gifts he had the gentle art of making enemies.

Pepys, Aubrey and Evelyn concur in the judgment that Petty was a most entertaining companion. Evelyn says he was a wonderful mimic. He could speak 'now like a grave orthodox divine; then falling into the Presbyterian way; then to Fanatical, to Quaker, to Monk, and to Friar and to Popish Priest'. The gift he exercised among his friends.

My Lord D. of Ormond once obtained it of him, and was almost ravished with admiration; but by and by he fell upon a serious reprimand of the faults and miscarriages of some Princes and Governors, which, though he named none, did so sensibly touch the Duke, who was then Lieutenant of Ireland, that he began to be very uneasy, and wished the spirit layed, which he had raised; for he was neither able to endure such truths, nor could but be delighted. At last he turned his discourse to a ridiculous subject, and come down from the joint-stool on which he had stood, but my lord would not have him preach any more (Evelyn).

My lord Duke was not the first or last person to fail to relish a joke against himself.

In *The Londoners* a challenged party names garden hoes as the weapons. That was Mr Robert Hichens's fun. In real life, Petty, challenged to mortal combat by a Cromwellian soldier, pleaded his myopia and demanded that the duel should take place in a cellar and the weapons be axes.

A man like this makes friends or at least admirers, also enemies. Long before the king enjoyed his own again, Petty had a host of enemies. When the king returned, one might have expected that Petty's position would be critical. According to his own account he *did* lose something, but he was knighted and the losses, such as they were, did not seem to stay the growth of his fortune. At the Restoration he was already prosperous and he died wealthy: Perhaps the

explanation is that Petty was really as great a public benefactor as he thought he was. Perhaps the reason is personal. King Charles loved wits (in the old and new sense of the word) and Petty was a wit. The scanty specimens of what Petty's modern representative calls 'Rabelaisian' printed from the Petty papers would not have appealed to such a connoisseur in this genre as the king—we know from Halifax that the king liked to be the raconteur in this field and indeed repeated himself often—but he would have relished a good mimic. Still more important might have been their common virtuosity.

Charles was interested in experimental science, and although Petty certainly knew more than the king, he may not have known very much more. Neither Charles nor James would have been able to find more common ground with Isaac Newton than in a later age Bonaparte found with Laplace. But the ingenious Dr Petty, who had resuscitated half-hanged Ann Green (which would be a capital story if well told), invented an unsinkable ship, had a dozen plans for doubling the king's revenue, and knew something of everything, probably did more than Wilkins to interest the king in the new society of virtuosos (how the king must have relished the story of the planting of horns in Goa*), and he may incidentally have interested the king in his business affairs. This is all speculation; what is sure is that when Petty was back in London and able to renew personal intercourse with John Graunt, their relation was no longer that of client and patron. For a few years more, Graunt was to be a solid merchant, but before long Petty was the patron and Graunt the client.

At this point it will be convenient to conclude the biographical facts relating to Graunt. I take them mainly from Aubrey.

Graunt continued to be a prosperous city tradesman for many years after his first meeting with Petty. 'He was', says Aubrey, 'a man generally beloved; a faithful friend. Often chosen for his prudence and justice to be an arbitrator; and he was a great peace-maker. He had an excellent working head, and was facetious and fluent in his conversation.' Pepys thought as well of Graunt as did Aubrey, admiring both his conversation and his collection of prints—'the best collection of anything almost that ever I saw'.

From the Restoration for several years Graunt figures in London intellectual society (he was elected F.R.S. in 1663), but a material calamity was at hand. The Fire of 1666 no doubt caused Graunt direct financial loss; this might have been repaired. But, although brought up in Puritan ways, 'he fell', to quote Aubrey, 'to buying and reading of the best Socinian bookes, and for severall

* Sir Philiberto Vernatti, Resident in Batavia, had certain inquiries sent him by order of the Royal Society. The eighth question was: 'What ground there may be for that Relation, concerning Horns taking root, and growing about Goa?' This is Sir Philiberto's answer: 'Inquiring about this, a friend laughed, and told me it was a jeer put upon the Portuguese, because the women of Goa are counted much given to lechery' (Sprat's *History of the Royal Society of London*, 2nd ed. London 1702, p. 161).

əars continued of that opinion. At least, about...he turned a Roman Catho-
ղue, of which religion he dyed a great zealot.'

Graunt's path to Rome was similar to that of young Edmund Gibbon, but ıe results on the career of a city tradesman in the days of Oates *triumphans* ere more serious than a visit to Lausanne. Graunt became bankrupt. His ame dropped out of the list of the Royal Society after 1666, and in 1674 he ied. There is evidence that in these last years of worldly misfortune, when the heel had come full circle since Graunt had secured the Gresham professorship ›r Petty, Petty helped Graunt. When Petty was in Ireland, Graunt acted in ›me sort as his London agent, and Petty conceived a plan of settling Graunt in :eland. But (we have, of course, only Petty's word for this) Graunt was not an ısy man to help; it is possible, of course, that he may have resented Petty's dmonitions. 'You have done amiss in sundry particulars, which I need not ıention because you yourself may easily conjecture my meanings. However we ›ave these things to God and be mindful of what is the sum of all religion, and f what is and ever was true religion all the world over.' This is an extract from letter of January 1673 to Graunt (*The Petty-Southwell Correspondence*, p. xxix) rinted by the late Marquis of Lansdowne. If Lord Lansdowne was right (the ʼhole letter is not printed) in thinking this a reference to Graunt's conversion ›r perversion) 'of which', says Lord Lansdowne, 'Petty seems to have dis- pproved on temporal rather than spiritual grounds', it might have hurt a ensitive man.

Graunt died on Easter Eve 1674 and was buried the Wednesday following in ›t Dunstan's church in Fleet Street. 'A great number of ingeniose persons ›ttended him to his grave. Among others, with teares, was that ingeniose great ʼirtuoso, Sir William Petty, his old and intimate acquaintance, who was sometime ı student of Brasenose College.' Sir William outlived his friend thirteen years ınd lies in Romsey Abbey. Until a descendant in the nineteenth century (the ›hird Marquis of Lansdowne) erected a monument, 'not even an inscription ndicated that the founder of political economy lay in Rumsey Abbey' (Fitz- naurice, p. 315).

Graunt had a son who died in Persia and a daughter who, according to Aubrey, became a nun at Ghent. Nothing is known of descendants.

Petty's widow was raised to the peerage and her elder sons, Charles and Henry, died without issue. But the title was revived in favour of the grandson of John Fitzmaurice, the second surviving son of Thomas Fitzmaurice, Earl of Kerry, who, as the above-mentioned grandson remarked, had 'married luckily for me and mine, a very ugly woman who brought into his family whatever degree of sense may have appeared in it, or whatever wealth is likely to remain in it'. This ill-favoured woman was Petty's daughter Anne, to whom her father wrote:

My pretty little Pusling and my daughter Ann
That shall bee a countesse, if her pappa can.

The cynical grandson was George III's prime minister and afterwards his *bête noire*, 'The Jesuit of Berkley Square' and first Marquis of Lansdowne.

Of the two friends, one has left an intellectual monument only; descendants of the other have been famous in English history.

Of these, best known are the first and third Marquises of Lansdowne, William (1737–1805) and Henry (1780–1863). Of the first marquis, much better known as Lord Shelburne (the title created for Lady Petty), every schoolboy—not only Macaulay's schoolboy—has heard; the quarrel between Charles Fox and Shelburne, the party split, the coalition ministry and so on. Schoolboys who have reached the sixth and Lecky's *History of England in the Eighteenth Century*, know a little more. Shelburne, who had much more than a tincture of his great-grandfather's ability and applied himself to economic studies, was one of the earliest to appreciate the importance of Adam Smith and was highly thought of by two good judges of scientific ability, Benjamin Franklin and Jeremy Bentham.

As a public man, no parliamentary statesman before or since obtained so universal a dislike, a positive hatred shared by those who knew him and those who did not.

> There is certainly nothing in the actions of Shelburne to justify this extreme unpopularity. Much of it was, I believe, simply due to an artificial, overstrained, and affectedly obsequious manner, but much also to certain faults of character, which it is not difficult to detect. Most of the portraits that were drawn of him concur in representing him as a harsh, cynical, and sarcastic judge of the motives of others; extremely suspicious; jealous and reserved in his dealings with his colleagues; accustomed to pursue tenaciously ends of his own, which he did not frankly communicate, and frequently passing from a language of great superciliousness and arrogance to a strain of profuse flattery (Lecky, **5**, 136).

How far some of these characteristics may be recognized in Shelburne's ancestor, we shall inquire in due course.

The contrast between Malagrida* and his son Henry is shattering. It is *this* Marquis of Lansdowne of whom nearly everybody thinks when he sees the title in a book, and rightly so. Walter Bagehot wrote:

> You may observe that when an ancient liberal, Lord John Russell, or any of the essential sect, has done anything very queer, the last thing you would imagine anybody would dream of doing, and is attacked for it, he always answers boldly, 'Lord Lansdowne said I *might*'; or if it is a ponderous day, the eloquence runs, 'A noble friend with whom I have had the inestimable advantage of being associated from the commencement (the infantile period I might say) of my political life, and to whose advice,' etc., etc., etc.—and a very cheerful existence it must be for 'my noble friend' to be expected to justify—(for they never say it except they have done something very odd)—and dignify every aberration. Still it must be a beautiful feeling to have a man like Lord John, to have a stiff, small man

* Malagrida was an Italian Jesuit settled in Portugal who was burned in 1761. The supposed jesuitical propensities of Shelburne led to the name becoming his popular title. Hence Goldsmith's unintended *mot*: 'Do you know that I never could conceive the reason why they call you Malagrida, for Malagrida was a very good sort of man.'

bowing down before you. And a good judge (Sydney Smith) certainly suggested the conferring of this authority. 'Why do they not talk over the virtues and excellencies of Lansdowne? There is no man who performs the duties of life better, or fills a high station in a more becoming manner. He is full of knowledge, and eager for its acquisition. His remarkable politeness is the result of good nature, regulated by good sense. He looks for talents and qualities among all ranks of men, and adds them to his stock of society, as a botanist does his plants; and while other aristocrats are yawning among stars and garters, Lansdowne is refreshing his soul with the fancy and genius which he has found in odd places, and gathered to the marbles and pictures of his palace. Then he is an honest politician, a wise statesman, and has a philosophic mind', etc., etc. Here is devotion for a carping critic; and who ever heard before of *bonhomie* in an idol? (Bagehot, *Works*, **2**, 64–5).

Of the father, Atticus (an alias of 'Junius') wrote:

The Earl of Shelburne had initiated himself in business by carrying messages between the Earl of Bute and Mr. Fox, and was for some time a favourite with both. Before he was an ensign he thought himself fit to be a general, and to be a leading minister before he ever saw a public office. The life of this young man is a satire on mankind. The treachery which deserts a friend, might be a virtue compared to the fawning baseness which attaches itself to a declared enemy (*Letters of Junius*, Wade's edition, **2**, 248).

Naturally justice was no more to be expected in eighteenth-century newspaper diatribes than in the twentieth century, but a clever caricaturist does not represent Charles Fox as a living skeleton. Those who attacked the son—there were such people—took a different line, as Bagehot hints. Perhaps even in his very different character something of the ancestral Petty survives. We shall try to discover what this was.

Forty years ago Hull brought out an edition of Petty's tracts in which he included Graunt's work. In 1927 the fifth Marquis of Lansdowne printed a selection from the Petty papers and in 1928 the correspondence between Petty and his wife's cousin,* Sir Robert Southwell (*The Petty-Southwell Correspondence*, edited by the Marquis of Lansdowne, London 1928).

We shall have to examine in detail both the 'works' and the 'papers', but, as a light upon the character of Petty, the Southwell correspondence is the strongest we have. Southwell himself was some generations farther away from adventuring than Petty. He came of an 'undertaker' stock—the adventurers in Ireland of Queen Elizabeth's time—and his father was vice-admiral of Munster before him. He was born in 1635 (died in 1702), regularly educated (Queen's College, Oxford and Lincoln's Inn), knighted in 1665, for some time Clerk of the Privy Council, in the diplomatic service, held other offices, was a member of parliament and eventually settled in a country house near Bath. He was President of the Royal Society 1690–5. He might be described as a lesser William Temple; better educated and less selfish, not so able, but with the same cool, cautious judgment; a psychological antithesis of his correspondent.

* Petty married in 1667 Lady Fenton, widow of Sir Maurice Fenton and daughter of Sir Hardress Waller who, knighted in 1629, fought for the Parliament and was one of the King's judges; he was a major general in Ireland in 1650–1 and a patron of Petty there.

The correspondence covers the eleven years 1676–87. Both men were, even by modern standards, middle aged. They write one to another with complete frankness; there is a remarkable absence of the elaborate verbal formalities which in seventeenth-century and even eighteenth-century letters are so wearisome.

Petty's side of the correspondence consists roughly of domesticities 10 parts, eager accounts of his quarrels and law suits concerning money 40 parts, discussion of papers or projected papers 40 parts, add autobiographical boasting to make up the 100.

In the purely domestic part of the correspondence, Petty is seen as a kind, good-natured father interested in the doings of his relations by marriage, also as a very bad judge of others' feelings. I remember to have read an unpublished letter by the famous Edwin Chadwick, the great and very unpopular sanitarian of a century ago. It was written to a friend whose wife had just died of puerperal fever. Chadwick expressed regret in the shortest possible formula and assured his correspondent that the best solace he could have would be to assist in pushing forward a bill (which I think he enclosed) to promote some sanitary reform which would have the effect of making it less likely that other men would lose their wives in childbed. I remember thinking that, however sensible the recommendation, the man who gave it was not likely to bring much comfort to his friend.

Petty was very much like Chadwick here. Southwell lost his wife in 1681 and Petty condoled with him as follows:

> When your good father dyed, I told you that hee was full of years and ripe fruit, and that you had no reason to wish him longer in the paines of this world. But I cannot use the same Argument in this Case for your Lady is taken away somewhat within half the ordinary age of Man and soon after you have been perfectly married to her; for I cannot believe your perfect union and assimulacon was made till many years after the Ceremonies at Kinsington.
>
> What I have hitherto said tends to aggravate rather than mitigate your sorrow. But as the sun shining strongly upon burning Coles doth quench them, so perhaps the sadder Sentiments that I beget in you may extinguish those which now afflict you. The next Thing I shall say is, That when I myself married, I was scarce a year younger then you are now, and consequently do apprehend That you have a second Crop of Contentment and as much yet to come as ever I have had.

This remark, curiously enough, was not well received.

> You doe not onely condole the great loss I have sustained in a wife, but you seeme to think it reparable....But when by 19 yeares conversation I knew the greate vertues of her mind, and discover since her death a more secrett correspondence with Heaven in Acts of Pietye and devotion (which before I knew not of), you will allow me, at least for my Children's sake, to lament that they have too early lost their guide.

Petty could not, it seems, understand that Southwell was wounded and returned to the charge in a letter which is lost. That letter provoked a reply

which even Petty could not misunderstand and elicited an apology (*Correspondence*, p. 90).

Petty was quite incorrigible. A few years later Southwell had another family bereavement and is condoled with in the following terms:

> That by the death of your Father, Mother and Sister, of Sir Edward Deering and your three nephews, you are the Head and Governor of both Familyes. That by the death of Rupe, Ingenious Neddy culminates; and by that of your Excellent Lady you are entitled to that million I mentioned of unmarryed teeming Ladyes.

Once again, Southwell was not comforted. 'Cousin, you doe wipe off Teares at a very strange rate, but why did nature furnish Them if there must be no Sorrow?'

Petty had a very quick perception of when and where *his* shoe pinched, but no imaginative sympathy.

Passing to Petty's financial affairs and lawsuits, the position was this. By original grants, by purchase and in various ways, Petty had widely scattered Irish interests. Questions of the validity of the original grants, of rent charges due to the crown or to other grantees, of matters of fact and matters of law were endless. Petty saw himself steadily as a great public benefactor harassed by scoundrels, and it never occurred to him even as a theoretical possibility that others had rights. Of his manner of proceeding the editor of the correspondence gives a typical example (*Correspondence*, p. 90). In 1681 Petty gave evidence before Lord Chief Baron Hen as to 'Soldier's land' which he had bought in Kerry and, it seems, the court decided against him.

> Petty gave vent to his chagrin in a long and scurrilous lampoon against the offending judge, entitled: 'HENEALOGIE or the legend of Hen-Hene and Pen-Hene', in two parts. Whereof the first doth in 24 chapters of Raillery, contain the enchantements, metamorphoses and merry conceits relating to them. The second part contayning (in good earnest) the foolish, erroneous, absurd, malicious and ridiculous 'JUDGEMENTS of HEN-HENE'. Fortunately perhaps for the repute of its author, this diatribe was never made public.

Fortunately, also, for a more material reason; it would probably have led to a *second* incarceration for contempt of court.

Southwell evidently viewed his good cousin's proceedings with a mixture of gentlemanlike annoyance and practical minded contempt. He expressed these feelings more than once; the following extract from a letter of 1677 is typical; the particular suit in progress to which reference is made was a claim for £5000 in respect of a sum of £2500 actually advanced by Petty to the Farmers of Revenue.

> And suffer from me this expostulation, who wish your prosperity as much as any man living; and having opportunities to see and heare what the temper of the world is towards you, I cannot but wish you well in Port, or rather upon the firm Land, and to have very little or nothing at all left to the mercy and good will of others. For there is generally imbibed such an opinion and dread of your superiority and reach over other men in the wayes of dealing, that they hate what they feare, and find wayes to make him feare that is

feard. I doe the more freely open my soul to you in this matter, because tis not for the vitells that you contend, but for outward Limbs and accessions, without which you can subsist with Plenty and Honour. And therefore to throw what you have quite away, or at least to put it in dayly hazard onely to make it a little more than it is, Is what you would condemne a thousand times over in another. And you would not think the Reply sufficient that there was plain Right in the Cause and Justice of their side, for iniquities will abound and the world will never be reformed.

After all this is said, I mean not that you should relinquish the pursute of your 2500£, which is money out of your Pockett and for which you are a Debtor unto your Family. But for other pretensions, lett them goe for Heaven's Sake, as you would a hott coale out of your hand: and strive to retire to your home in this Place, where you had the respect of all, and as much quiet as could be in this life, before your medling with that pernicious business of the Farme.

There is no reason to suppose that Petty ever took such sensible advice. Yet, somehow, he kept his head well above water.

In the later part of the correspondence Petty indulges in that complacent financial retrospect which he inserted in his Will and I have, perhaps too harshly, described as autobiographical boasting. It is possible that Southwell had heard of these financial triumphs rather often; at least there is a hint of this in the following:

I will onely note that since you are soe Indulgent as to think me worthy of being your Depositary in this great Audit, and expect by the Course of Nature that I should speake when you are Silent, you must allow me liberty without blame to aske questions when you seeme defitient or Redundant.

That you are defitient may be suggested when, on the fortunate syde, I find noe Item for my Lady or of the hopefull stock she has brought you (p. 227).

The shrewd thrust of the last sentence was deadly. The subject does not recur.

I have indicated the character of the non-scientific part of the correspondence because we must examine Petty's scientific writings in greater detail. I think, however, we have enough to justify a provisional diagnosis of Petty's psychological type.

In literature and in life the perennial boy is often encountered. But while Peter Pan and Mr Reginald Fortune make far more friends than foes, that is not so true of their living counterparts. The exuberant flow of ideas and schemes, the intense and restless interest in *everything* which is characteristic of the clever child, often is extraordinarily attractive when it is associated with and controlled by the trained intelligence of a man. But the bad as well as the good points of a childlike or adolescent soul* are to be brought into the account. The

* The first Marquis of Halifax said of King Charles that 'his inclinations to love were the effects of health and a good constitution, with as little mixture of the *seraphic* part as ever man had', and Petty held that the King was typical. In *The Petty Papers* (no. 93 of vol. 2) there is a memorandum headed 'Californian Marriages with the Reasons thereof'. 'In California', says Petty, '6 men were conjugerted to 6 women in order to beget many and well conditioned children, and for the greatest venereall pleasure, in manner following, viz.'

He then sets out the plan. One man 'excelling in strength, nimbleness, beauty, wit, courage

clever child is often naïvely and intensely selfish, and so remains as the eternal boy; his quite crude and unashamed egoism, his inability to understand that others have feelings and even rights, repel as strongly as his intellectual freshness attracts. How far he is a success in life depends on which way the balance turns.

Petty seems to me a good example of this psychological type; its good points, the restless energy and exuberant flow of ideas, were sources of strength in such a time as that of the Civil War and Restoration, which, particularly the Restoration period, was in virtues and vices an age of grown-up children. Indeed his emotional adolescence may have shielded him from the deadly enmity of real men. Its bad points made him enemies, but they were children like himself. Nearly a century later, in a time of adults, these same characteristics, restless intellectual energy and vanity, exhibited by one no longer a rollicking adventurer but a great landowner, produced an unfavourable balance and we have 'Malagrida'. In Malagrida's son, one has a change; the attractive traits, the eager interest in all sorts of things is still there, but the childish hungry vanity has been softened or sublimed. The cynic may say that it was easy for a great Whig lord 150 years ago to be agreeable, to keep himself *hors concours*; perhaps it was, although the *Dropmore Papers* raise doubts. The fact, however, is certain. In the third Lord Lansdowne one sees the good and in the first the bad effects of the perennial boyishness of the ancestor. The ancestor lived in a state of society where the good points outweighted the bad points. That is why, although he made enemies and was often vexed, he was able to view his career with complacency and to bequeath a great fortune. But it is not Petty as a man but Petty as a scientific worker who is the proper object of my study.

How far does the psychological make-up which, as I think, characterized Petty conduce to scientific investigation? We might expect that it would be an immense stimulus to pioneering, that such a man would direct attention to a number of problems which deserved study, but that it would not lead to the production of any solid contribution to knowledge. Our task is to examine in some detail Petty's scientific work.

and good sense' subsequently called the Hero, is allowed four women for his sole use. One Great Rich Woman is allowed five men who are to serve her when she pleases, but another woman is allotted to the five men for use in common by the five.

It may be said this fable is only an after dinner jest—perhaps that *is* the whole explanation. But Petty does go to the trouble of financial calculations, and does seem to suggest a serious consideration. ('The encrease of children will be great and good.' 'No controversy about joynture, dower, maintenance, portion etc.') Nobody emotionally adult would be likely to make Californian Marriages a basis for practical statecraft.

II. PETTY'S SCIENTIFIC WORK

It is no part of my undertaking to survey the whole of Petty's scientific activities, but to speak only of his medical and vital statistical work.

In Hull's edition of Petty's writings, the editor discusses Petty's status as an economist and remarks that Petty's view that value depended upon labour was probably derived from Hobbes. The corn rent of agricultural lands was in Petty's view determined by the excess of their produce over the expenses of cultivation, paid in corn, and the money value of the excess will be measured by the amount of silver which a miner, working for the same time as the cultivator of the corn land, will have left after meeting his expenses with a part of the silver he secures (Hull, p. lxxiii). Why there should be any surplus, he explains by density of population.

Prof. Hull refrained from attempting to assess Petty's work in terms of modern economic theory. A mere medical statistician will naturally follow this example. More than a century ago, Mr Chainmail had learned from Mr MacQuedy that the essence of a safe and economical currency was an interminable series of broken promises and added: 'There seems to be a difference among the learned as to the way in which the promises ought to be broken; but I am not deep enough in their casuistry to enter into such nice distinctions.' Medical statisticians may well adopt Mr Chainmail's modest attitude towards the whole field of economic theory. Confining ourselves to statistics, we must consider what Petty thought should be done and what he actually did himself.

Under the first heading, praise can be unstinted. More than 150 years before the establishment of the General Register Office, Petty specifically proposed the organization of a central statistical department the scope of which was wider than that of our existing General Register Office. It was to deal not only with births, marriages, burials, houses, the ages, sexes and occupations of the people, but with statistics of revenue, education and trade (see *The Petty Papers*, 1, 171–2). He did not confine himself to vague recommendations, but drew up an enumeration schedule to be used for each parish. On this was to be entered: The number of housekeepers and of houses; the number of hearths; the number of statute acres; the number of people by sex and in age groups, viz. under 10, between 10 and 70, over 70; for males those aged 16 to 60, and for females those between 16 and 48 and how many of these latter were married; how many persons were incurable impotents and how many lived upon alms. This, it will be noted, is a better enumeration schedule than any used in England before the census of 1821. Further in his notes (printed in *The Petty Papers*) are various suggestions for the utilization of data collected in this way.

The most striking is this: 'The numbers of people that are of every yeare old from one to 100, and the number of them that dye at every such yeare's age, do shew to how many yeare's value the life of any person of any age is equivalent

and consequently makes a Par between the value of Estates for life and for years' (*The Petty Papers*, **1**, 193).

This is, I think, the most remarkable thing Petty ever wrote, for it *suggests* that he had grasped the principle of an accurate life table, viz. a survivorship table based upon a knowledge of rates or mortality in age groups. No such table was constructed from population data until the end of the eighteenth century, because until then data of the age distribution of the *living* population were not obtained. Whether Petty also realized that under certain conditions a life table could be constructed without knowledge of the ages of the living population is a controversial matter which I shall discuss later on.

Then he makes suggestions which are relevant enough to modern demographic problems.

> By the proportion between marriages and births, and of mothers to births, may be learnt what hindrance abortions and long suckling of children is to the speedier propagation of mankind; as also the difference of soyles and ayres to this foecundity of women.
>
> By the proportion between maryd and unmaryd teeming women, may be found in what number of yeeres the present stock of people may bee encreased to any number assigned answerable to the defect of the peopling of the nation for strength or trade.

There are not wanting some suggestions which imply that even if Petty's opinion of the Faculty were higher than that of Sydenham (whom we honoured *posthumously*) it was tinged with scepticism.

> Whether they [viz. fellows and licentiates of the College of Physicians] take as much medicine and remedies as the like number of any other society.
>
> Whether of 1000 patients to the best physicians, aged of any decade, there do not die as many as out of the inhabitants of places where there dwell no physicians.
>
> Whether of 100 sick of acute diseases who use physicians, as many die and in misery, as where no art is used, or only chance. (*The Petty Papers*, **2**, 169–70.)

This statistical experiment has not yet been performed and indeed might be hardly so conclusive as Petty implied.

When one passes from what Petty suggested to what he actually did himself, our praise must be qualified. As Prof. Hull said, he was 'more than once misled into fancying that his conclusions were accurate because their form was definite'.

In judging Petty it is but fair to contrast him with College contemporaries whose names are more honoured by us. Among his contemporaries in the College were Thomas Browne and Thomas Sydenham. Browne was a much older man than Petty, Sydenham almost his coeval. Of Browne's quality as a physician we know nothing; but his literary influence indirectly—through Samuel Johnson—and directly upon generations of readers has been greater than that of any other practising medical man. Browne, like Petty, had an enormous range of interests and his book learning was greater. But, as we shall see, when

he tackles a problem of demography, Petty's rashest guesses seem by comparison as soberly scientific as an annual report of the Registrar-General.

Sydenham was an iconoclast in clinical practice and believed himself to be emancipated from the rule of ancient authority. No fantastic arithmetical calculations are to be found in *his* writings. In fact, with a single exception (*Observations Medicae*, 2, i), no arithmetic at all. It never seems to have entered his mind, although his greatest work purports to give the history of the diseases in London through a generation, that the arithmetical statements of the London Bills of Mortality were of any value whatever.

Sydenham was too wise a man for us to think that he rejected the evidence because the data were compiled by illiterate old women. He would have known that the sworn searchers had the loquacity of their sex and rank and were likely to ask what 'the doctor said'. He rejected it, because counting and measuring things did not come within his purview, just as the first beginnings of pathology and medical chemistry seemed to him irrelevant.

For the most part, Petty's statistical work was severely practical, but there is one excursion into theory which is interesting. It is to be found in a section of his tract on the use of what he calls Duplicate Proportion and is reprinted by Hull (pp. 622–3).

Petty states that there are more persons living between the ages of 16 and 26 than in any other decade of life. The statement is not true for modern populations and was probably not true for the English population of Petty's time. In 1861–71 (before the fall in the birth rate and infant mortality rate) there were 5·4 millions living under 10, and 4·0 between 15 and 25). But perhaps Petty meant that there were more living in the decade 16 to 26 than in any *later* decade, in which case his statement was of course right unless the birth rate was falling.

He then asserts that the

> Roots of every number of Men's Ages under 16 (whose Root is 4) compared with the said number 4, doth show the proportion of the likelyhood of such men reaching 70 years of Age. As for example: 'Tis 4 times more likely that one of 16 years old should live to 70, than a new born Babe. 'Tis three times more likely, that one of 9 years old should attain the age of 70, than the said infant. Moreover, 'tis twice as likely, that one of 16 should reach that Age, as that one of four years old should do it; and one third more likely, than for one of nine.

We have no life table for England in 1674. Perhaps the nearest modern experience might be the Liverpool Table calculated by Farr seventy years ago. According to that table the chance of a new-born child living to be 65 was 0·0976 and the chance of a person of 15 living to 65 was 0·202, which is about double the infant's chance, not four times as large. For the Healthy Districts, the chances are 0·4246 and 0·54585; that is, in a ratio of 1·28 to 1.

Petty's statements are wildly wrong. The interesting point is how did he reach them? The only figures he had were printed by Graunt.

This 'Life Table' gives l_x as follows:

l_0	100	l_{46}	10
l_6	64	l_{56}	6
l_{16}	40	l_{66}	3
l_{26}	25	l_{76}	1
l_{36}	16		

Now if we take 2 as the survivors to 70 (it does not of course matter what the numerator is for comparative purposes), then the infant's chance of surviving to 70 is 0·02 and the person of 16 has the chance $1/20 = 0{\cdot}05$, a ratio of 2·5, not wildly different from the Liverpool Table figure and very different from 4·0.

A fortiori when Petty, having passed above age 16, asserts that 'it is five to four, that one of 26 years old will die before one of 16; and 6 to 5 that one of 36 will die before one of 26', we are in a region of pure fantasy because, even if he had had the statistical data, Petty would not have had the technical knowledge to solve the problem involved, viz. to find the probability that of two lives aged respectively x and y, the former will fall before the latter.

If we keep within the range of the simple arithmetic which Petty used, the result cannot be obtained.

He then passes to this statement:

> To prove all which I can produce the accompts of every Man, Woman, and Child, within a certain Parish of above 330 Souls; all which particular Ages being cast up, and added together, and the Sum divided by the whole number of Souls, made the Quotient between 15 and 16; which I call (if it be Constant or Uniform) the Age of that Parish, or *Numerus Index* of Longaevity there. Many of which Indexes for several times and places, would make a useful Scale of Salubrity for those places, and a better Judg of Ayers than the conjectural Notions we commonly read and talk of. And such a Scale the *King* might as easily make for all his Dominions, as I did for this one Parish.

The puzzle is to discover why Petty thought this statistical experiment proved his point and why he regarded the mean age of the population of a parish its index of longevity. The first question I cannot answer at all; about the second I can make a guess. *If* the parish population were supported solely by births and there was no migration, then, if the death rates at ages did not vary, the population would be a stationary population and both the mean age of the living and the mean age at death would be constant. The expectation of life is greater than the mean age of the living unless the rates of mortality at early ages are very high and the more favourable the rates of mortality the greater will be the difference. In Petty's day, when mortality at early ages was very high, the two constants were probably not far apart, but it is certain that both expectation of life and mean age of a life table population were greater than 16; probably of order 28 to 32.

I think we may be sure that the parish Petty counted was not stationary in the statistical sense, but had an excess of births over deaths, and that his average threw no light upon the rates of mortality.

Passing to practical statistics, it will be convenient first to note rapidly statistical observations which are incidental in treatises of primarily financial or economic interest. In the *Verbum sapienti,* which although not printed until 1691 was written as early as 1665, Petty attempts to reckon what a man is worth. Here is the method. He concludes from financial data that the annual proceed of the Stock or Wealth of the nation yields 15 millions, but that the expenses of the nation are 40 millions. So the balance of 25 millions must be derived from the labour of the people. He assumes that the population is 6 millions and that half of these can work, and earn £8. 6*s.* 8*d.* a head per annum. This would be 7*d.* a day, abating 52 Sundays and half as many other days for sickness, holidays, etc. 'Whereas the Stock of Kingdom, yielding but 15 Millions of proceed, is worth 250 Millions; then the People who yield 25, are worth 416 2/3 Millions. For although the Individuums of Mankind be reckoned at about 8 years purchase; the Species of them is worth as many as Land, being in its nature as perpetual, for ought we know.'

Why an individual's working life is worth only 8 years' purchase is not clear. One would be inclined to put it as the average number of years lived in the working period of life. Perhaps Petty took Graunt's table and worked out the average number of years of life lived between the ages of 16 and 56; it *is* nearly 8.

He then calculates the money loss due to 100,000 dying of the plague and makes it nearly 7 millions, adding that £70,000 would have been well disposed in preventing this 'centuple loss'. Perhaps this is the first printed statement of the neglected truth that public health measures pay.

Since Petty's day, others, including Farr himself, have done sums of this kind; it is a popular occupation in the United States of America.

Farr went to work more elaborately, making out a balance sheet of a man from the cradle to the grave. But the principle was much the same. We cannot say it is a *wholly* useless pastime. There is of course the difficulty that if more lives are saved the price of labour might fall. But to Petty that would have been no difficulty, because he held that wealth is *purely* relative, viz. that if the income of each person in a community is halved, everybody is as well off as before.

In the *Political Anatomy of Ireland,* Petty seeks to determine war losses in Ireland.

The number of the People being now *Anno* 1672 about 1,100,000 and *Anno* 1652 about 850 M. Because I conceive that 80 M. of them have in 20 years encreased by generation 70 M. by return of banished and expelled *English*; as also by the access of new ones, 80 M. of New *Scots,* and 20 M. of returned *Irish,* being all 250 M.

Now if it could be known what number of people were in *Ireland* Ann. 1641, then the difference between the said number, and 850, adding unto it the increase by generation in 11 years will shew the destruction of people made by the Wars, *viz.* by the Sword, Plague and Famine occasioned thereby.

I find by comparing superfluous and spare Oxen, Sheep, Butter and Beef that there was exported above 1/3 more *Ann.* 1664 than in 1641, which shews there were 1/3 more of

people, *viz.* 1,466,000. Out of which Sum take what were left Ann. 1652, there will remain 616,000 destroyed by the Rebellion.

Whereas the present proportion of the *British* is as 3 to 11; But before the Wars the proportion was less, *viz.* as 2 to 11 and then it follows that the number of *British* slain in 11 years was 112 thousand Souls; of which I guess 2/3 to have perished by War, Plague and Famine. So as it follows that 37,000 were massacred in the first year of Tumults: So as those who think 154,000 were so destroyed, ought to review the grounds of their Opinions.

It follows also, that about 504 M. of the *Irish* perished, and were wasted by the Sword, Plague and Famine, Hardship and Banishment, between the 23 of *October* 1641 and the same day 1652. Wherefore those who say, That not 1/8 of them remained at the end of the Wars, must also review their opinions; there being by this Computation near 2/3 of them; which Opinion I also submit.

Assuming, which is rash, that the estimates of population in 1672 and 1652 are correct, the assumption that population varied inversely as exportation of cattle seems bold. Might it not be that shipping facilities were better in 1664 than in 1641? Had there been no exportation we could not infer the population to be infinite.

Again Petty has multiplied the estimate for 1672 by 1·333. But he needed the population of 1664, which presumably was smaller than that of 1672. If his estimate is right, the population was increasing at the rate of about 12·5 thousands per annum, so he should have multiplied 1,000,000 not 1,100,000 by 1·333 and has overestimated the 1641 population by 133,330, and therefore the number destroyed by the same amount, an overstatement of 20 %. But this is not all. If we assign the decrement of population between 1652 and 1641 wholly to sword, plague and famine, we must assume that births continued at the peace-time rate; not a likely assumption. Lastly, it seems unreasonable to assign the casualties to the two races in precise proportion to their estimated numerical strength in the population of 1641.

How it follows that 37,000 were massacred in the first year of tumults I do not know.

In a later work (*Treatise of Ireland,* pp. 610–11) Petty has another shot at this problem.

He now assumes that Graunt's deduction from a Hampshire parish register, viz. that christenings are to burials in the ratio of 5 to 4, applies to Ireland, and that the death rate is 1 in 30, i.e. about what Graunt estimated for London and much higher than his estimate for the country. He then proceeds in this way. He estimates the population of 1653 to be 900,000 and that of 1687, 1,300,000. Then taking 1/30 for the death rate and 1/24 for birth rate, he makes the population of 1652, 985,000. He does not comment on the great decrease between 1652 and 1653; but there was still war in Ireland in 1652.

He now says that the population of 1641 was greater than that of 1687, 'as appears by the Exportations, Importations, Tyths, Grist-Mills and the Judgment of Intelligent Persons'. This time he takes the population to be 1,400,000—a little less than in the earlier estimate—and by the same kind of reasoning

again makes the war losses to be about 600,000. One is reminded of Hull's remark that Petty confused the accurate with the definite. Also one notes the inevitable tendency of a polemical writer—which Petty very decidedly was—to maintain his original assertion. Those of us who have *never* yielded to this temptation may cast stones at him. It is not I believe too cynical to say that *any* calculation Petty made would have made the war losses around 600,000.

Returning to the *Political Anatomy of Ireland*, we find here a distinct claim that the mean age at death (not the mean age of the living) measures longevity.

As to Longaevity, inquiry must be made into some good old Register of (suppose) 20 persons, who were all born and buried in the same Parish, and having cast up the time which they all lived as one man, the Total divided by 20 is the life of each one with another; which compared with the like Observation in several other places, will show the difference of Longaevity, due allowance being made for extraordinary contingencies and Epidemical Diseases happening respectively within the period of each Observation (p. 172).

Apart from what we should think the absurdity of basing important conclusions upon an average of 20—and Petty only gives 20 as a figure—the mean ages at death of different populations are not comparable unless in each place the population is stationary in the sense described above. But, since so acute a man as Edwin Chadwick made the same mistake in the nineteenth century as Petty in the seventeenth century and it continues to be made in various places in the twentieth century, we need not be superior.

We now come to Petty's purely statistical work which is concerned with the growth of population; before examining this in detail, it will be convenient to consider the methods available in the seventeenth century for estimating population and notions then current on what may be called the theory of population growth.

It is hard to believe that in the ancient world nobody studied demography arithmetically. There is evidence that the Romans enumerated citizens—the word census is pure Latin—and it has been suggested that the Romans made life tables. Gouraud, cited by Todhunter (*History of the Mathematical Theory of Probability*, p. 14), refers to a passage cited from Ulpian in the *Digest* which I have discussed elsewhere.* The question was of the value of annuities and the conclusion I reached was that Ulpian had no vital statistical basis whatever for his figures, that he simply began with the capital value the law gave for *any* usufruct and then, realizing that people do die eventually, made some subtractions, ending with the absurd (vital-statistically speaking) conclusion that after the age of 60 the rate of mortality was independent of age.

There is not, I think, any reason to believe that the practical Romans had anticipated Graunt and Petty.

That is not to say that nobody studied any demographical problems arithmetically. Indeed one fellow of the College of Physicians who has had—and will

* *Journ. Roy. Stat. Soc.* **103** (1940), 246.

continue to have—a hundred readers for every one reader of Graunt and Petty made an elaborate demographical calculation. This was Sir Thomas Browne. Sir Thomas devoted the sixth chapter of the sixth book of *Pseudodoxia* to the vulgar opinion that the earth was slenderly peopled before the Flood.

This vulgar opinion Sir Thomas found to be very wide of the mark. Indeed, far from the earth being slenderly peopled, 'we shall rather admire how the earth contained its inhabitants, than doubt its inhabitation: and might conceive the deluge not simply penall, but in some way also necessary, as many have conceived of translations, if *Adam* had not sinned, and the race of man had remained upon earth immortal'. Indeed Sir Thomas estimates that by the seventh century of the world's history its population amounted to 1,347,368,420. He reaches this result in the following way:

Having thus declared how powerfully the length of lives conduced unto populosity of those times, it will yet be easier acknowledged if we descend to particularities, and consider how many in seven hundred years might descend from one man; wherein considering the length of their dayes, we may conceive the greatest number to have been alive together. And this that no reasonable spirit may contradict, we will declare with manifest disadvantage; for whereas the duration of the world unto the flood was about 1,600 years, we will make our compute in less than half that time. Nor will we begin with the first man, but allow the earth to be provided of women fit for marriage the second or third first centuries; and will only take as granted, that they might beget children at sixty, and at an hundred years have twenty, allowing for that number forty years. Nor will we herein single out *Methuselah*, or account from the longest livers, but make choice of the shortest of any we find recorded in the Text, excepting *Enoch*: who after he had lived as many years as there be days in the year was translated at 365. And thus from one stock of seven hundred years, multiplying still by twenty, we shall find the product to be one thousand, three hundred forty seven millions, three hundred sixty eight thousand, four hundred and twenty.

	1.	20.
	2.	400.
	3.	8,000.
	4.	160,000.
Century.	5.	3,200,000.
	6.	64,000,000.
	7.	1,280,000,000.
		1,347,368,420.

Simply as a sum, there are difficulties about this result. If our 20 are equal numbers of males and females, it is not 20 which should be multiplied by 20 but 10. If they are all males, then women are left out of the reckoning. But, perhaps, as the Text does not record the ages of women, Sir Thomas esteemed them as ephemerids, sufficiently plentiful however to provide a wife for every husband. But then I think he should have said that the 20 to be begotten between 60 and 100 were all males. Anyhow the sum must be wrong because *some* of the 64,000,000 short-lived women of the sixth century should survive into the seventh. Indeed Sir Thomas uses his data a trifle capriciously.

We must surely play a game according to the rules. We are to accept the

Text word for word as it stands. But, omitting Adam, whose age at his begetting of Cain is not recorded, and Noah, who seems to have reached middle age—500 years—before becoming a father, the reproductive habits of eight fathers are recorded. Two begat males at the age of 65, one at 70, one at 90, one at 105, one at 162, one at 182 and one at 187. When this primary business was over, they are all recorded to have begotten an unspecified number of sons and daughters. So, if we are to be faithful to the Text, a very much more complicated arithmetical problem presents itself. A male begets another male at an average age of about 100, he then begets males and females at an unspecified rate for say another 600 years, required the law of increase. The Text does *not* authorize Sir Thomas to start pre-diluvian breeding at 65 or to stop it at 100. His 'manifest disadvantage' is breaking the rules of the game.

Further, the Text does not entitle him to predicate of the other males the lengths of days and procreative exploits of the recorded eight.

All this, it may be said, is breaking a butterfly upon the wheel. Nobody now takes the statistics of the Authorized Version literally. The point is that Sir Thomas Browne *did*, but used them improperly. As Lord Chesterfield said to a Garter King at Arms of his day who had not followed the rules of heraldry, 'You foolish man, you don't know your own foolish business'.

Petty did not tackle pre-diluvian demography, but he did try his hand at an estimate of the world's population after the flood, 'To justify the *Scriptures* and all other good *Histories* concerning the *Number* of the People in Ancient Time' (p. 465).

As Petty was not going to allow the population of ancient times to be greater than in the seventeenth century, but to make it increase regularly from the time of Noah's Ark, common sense saved him from fantastic figures, but not from physiological difficulties. The rules of the game obliged him to start with eight landed from the Ark, so he thought it best to make them increase and multiply very fast indeed at first and progressively more slowly. At first he doubled the population every ten years, but by the birth of Christ has brought the period up to 1000 years. But doubling every ten years (in the first century from the Flood) leads one into difficulties.

We can allow the possibility of the four pairs emerged from the Ark producing 8 offspring in ten years and so becoming 16 in year 10, without too great difficulty. But ten years later they must number 32 and this *is* a difficulty. If the fecundity of the first settlers remains the same they will contribute 8 more children, giving us a population of 24, the balance of 8 must come from the four couples of children all of whom must be under 20, and this *is* a little difficult.

But at least we may say that there is nothing wholly fantastic in Petty's procedure. Petty does belong to a different arithmetical world from that of Browne. Here we may leave purely speculative demography.

To estimate the people of an area without counting them, we must count

something which has a connexion with the number of the people. We may count the tax-payers, the houses, the burials, the christenings or the acreage under corn—all or any of these items vary with the number of people.

I wish to keep separate the discussions of Petty's and Graunt's statistical researches, but in the matter now to be examined Petty used some of Graunt's methods and results, so these must be considered.

Graunt used three methods of estimation. In the first place, he surmised that the number of child-bearing women in a community might be about double the number of annual births 'forasmuch as such women, one with another, have scarce more than one child in two years'. Then he surmised that families were twice as numerous as women of child-bearing age. His reasoning was that women between 16 and 76 might be twice as numerous as women between 16 and 40 or 20 and 44 (i.e. of child-bearing age), and he thought of a family as centred round a married couple. Finally, he thought that the average family would consist of eight persons, the husband and wife, three children and three servants or lodgers. So, starting with 12,000 christenings, which he thought a fair measure of annual births, he reaches 24,000 women of fertile age, then 48,000 families and lastly 384,000 persons.

It is quite certain that Graunt's estimate of an annual fertility rate of 500 per 1000 was an enormous overstatement. In London in 1851, the ratio of legitimate births to married women aged 15–45 was 251·8 per 1000. There is no reason to believe that nuptial fertility changed appreciably between 1660 and 1860. But an error of this kind would lead him to an understatement of families. Now, however, another error saves him. We cannot be so positive that eight to the family is a great overstatement as we can that the marital fertility was not 500 per 1000, but it is much higher than any nineteenth-century finding. Using *this* multiplier saves Graunt in this sense, that his quaint rule gives almost precisely the right answer for the population of London nearly 200 years after his time.

The legitimate births registered in London in 1851 were 75,097. This, according to Graunt's rule, is to be multiplied by 32. The result is 2,403,104. The enumerated population was 2,363,236; the conjecture is only 1·7 % out. *Sic me servavit Apollo.*

Graunt's next method was experimental and very briefly described. He counted the numbers of families in certain parishes within the walls and found that '3 out of 11 Families per annum have died'. He then multiplies the burials for the year (13,000) by 11/3, and proceeds as before.

Finally, he took Newcourt's map of London and

> guessed that in 100 Yards square there might be about 54 Families, supposing every House to be 20 Foot in the front: for on two sides of the said square there will be 100 Yards of Housing in each, and in the two other sides 80 each; in all 360 Yards: that is 54 Families in each square, of which there are 220 within the Walls, making in all 11880 Families within

the Walls. But forasmuch as there die within the Walls about 3200 per *Annum*, and in the whole 13,000, it follows that the Housing within the Walls is ¼ part of the whole, and consequently, that there are 47,520 Families in and about *London*, which agrees well enough with all my former computations (p. 386).

These conjectures led Graunt to think that the rate of mortality in London was about 1 in 32. In his first essay on the growth of London (pp. 458–75) Petty bases himself upon that estimate, and in the series of papers (pp. 505–44) this remains the fundamental method, but Petty allows himself to modify the multiplier, not altogether without suspicion of bias. At a quite early stage he had satisfied himself that London was the largest city in the world and much larger than Paris. This is the kind of argument. For the three years 1682–84, the average of burials in London was 22,337 and for Paris 19,887. So if the rates of mortality were the same, London was larger than Paris.* If the rate of mortality in Paris were higher than in London then the population of London must be larger still. According to Petty (*a*) a larger proportion of the Paris population died in hospital, (*b*) the mortality in hospital was heavier in Paris than in London. So it follows that the general death rate of Paris was higher.

That at *London* the *Hospitals* are better and more desirable than those of *Paris*, for that in the best at *Paris* there die 2 out of 15, whereas at *London* there die out of the worst scarce 2 of 16, and yet but a fiftieth part of the whole die out of the *Hospitals* at *London*, and 2/5 or 20 times that proportion die out of the *Paris Hospitals* which are of the same kind; that is to say, the number of those at *London* who chuse to lie sick in *Hospitals* rather than in their own Houses, are to the like People of *Paris* as one to twenty; which shows the greater Poverty or want of Means in the People of *Paris* than those of *London*. We infer from the premisses, viz. the dying scarce 2 of 16 out of the *London Hospitals*, and about 2 of 15 in the best of *Paris* (to say nothing of l'*hostel Dieu*) that either the *Physicians* and *Chirurgeons* of London are better than those of *Paris*, or that the *Air* of *London* is more wholesome (p. 508).

These, however, are only logical deductions if the user of the hospitals in London and Paris is identical. If, as implied in the first part of the quotation, we think of hospitals in the sense which our elder contemporaries think of the old-fashioned poor law infirmaries, viz. as refuges for the sick poor, it would mean that in Paris more of the aged indigent died in institutions than in London and heavy mortality might well have nothing to do with the skill or lack of skill of the medical staff. If we think of hospitals in the modern sense, then heavy mortality might be a mere reflection of the resort to these hospitals of persons suffering from illnesses which needed special treatment. In any case, Petty can hardly have it both ways. In another essay (pp. 510–11) he contrasts the higher ratio of deaths to admissions at l'hostel Dieu of Paris with that of la Charité, argues that the excess in l'hostel Dieu is unnecessary and proceeds to calculate

* It should be remembered that the London of Petty's calculations is the whole area within the Bills. The calculations of Graunt described above did not include Westminster or the six out-parishes of Surrey and Middlesex which were within the Bills: Islington, Lambeth, Stepney, Newington, Hackney, Redriff.

what the French nation would gain by saving this excess. But he has not inquired whether the patients of the two institutions were *in pari materia.*

Here is an historical problem which might be solved by those familiar with the literature of the period. Its discussion would not be relevant here. It is, however, only just to Petty to say that, unless conditions deteriorated seriously in the following century, his strictures on l'hostel Dieu were justified. In Franklin's work (*La Vie Privée d'autrefois. L'Hygiène* (Paris 1890), pp. 177 *et seq.*) an appalling account of this hospital from the pen of the eminent surgeon Tenon, printed in 1788, is quoted. Tenon's description of the routine of this great hospital compares, unfavourably, with the story of the wounded in the Mesopotamian campaign which horrified England in the war of 1914–18. He remarks, *inter alia*, 'on ne guérissoit point de trépanés autrefois à l'Hôtel-Dieu, comme on n'en guérit pas encore aujourd'hui', and cites a court surgeon of the time of Louis XIV, i.e. a contemporary of Petty, to that effect. His account of the treatment of lying-in women is grotesquely horrible.

In another essay (pp. 533–6) Petty discusses methods of estimation more carefully than in his other papers.

He proposes to show that the population of London (within the Bills) in or about 1685 was approximately 696,000.

There are, he says, three methods: (1) From houses and families. (2) From an estimated death rate. (3) From the ratio of those who die of the plague to those who escape.

This last we may deal with at once. Petty asserts that Graunt had proved that one-fifth of the people died of the plague. But in 1665, 98,000 died of the plague; therefore the population was 490,000, and allowing an increase of one-third between 1665 and 1686 we reach 653,000.

Graunt could not have proved that one-fifth of the population died of the plague unless he knew what the population was, and he never claimed to have done so.

The other methods (which Graunt used) are rational.

To estimate houses, Petty used three methods. He says that in the Fire of 1666, 13,200 houses were burned and that deaths from these houses were one-fifth of total deaths, so he reckons the houses to have been 66,000. Then as burials in 1686 were to burials in 1666 as 4 to 3, he makes the houses of 1686, 88,000. He does not, however, say upon what basis the estimate of one-fifth of the deaths in 1666 stands.

Next, he gives an estimate of the houses in 1682 given him by those employed upon a map said to have been made in that year. This map has not been identified.

Lastly, he uses the return of hearths. In Dublin in 1685 the hearths were 29,325 and the houses 6400. In London the hearths were 388,000; so the houses on the Dublin ratio should be 87,000. In Bristol he says there were 5307 houses

and 16,752 hearths, which give 123,000 houses for London; the mean of the calculations is 105,000. The Hearth Office itself, he says, certified the number to be 105,315. He must now have a multiplier. He accepts Graunt's multiplier of 8 as valid for tradesmen's families, but allows for smaller families among the poor and larger among the rich, finally choosing 6. He then allows for double families in houses by adding 10,531 to his 105,315, and multiplying the sum by 6 has 695,076 for the population.

Petty's second way was from an estimated death rate.

Petty multiplies the average of the burials in 1684 and 1685 (23,212) by 30, which makes the population 696,360.

He now essays to prove that the death rate in *London* was 1 in 30. He uses four arguments, of which only one is strictly to the point, viz. Graunt's direct observation that three deaths occur annually in eleven families—which however involves the assumption of eight persons to the families observed. Two others are relevant, viz. observations, apparently direct, that in 'healthful places' the mortality is 1 in 50 and in nine country parishes 1 in 37. The fourth partly rests upon a statement which Graunt did *not* make, viz. that one of 20 children under 10 dies annually. This fictitious value Petty averages with the statement of a M. Auzout to the effect that the rate of mortality of adults in *Rome* is 1 in 40. It will be clear that Petty has proved nothing at all. What he has done is to make it unlikely that the rate of mortality was less than 1 in 30. That, perhaps, was enough. One has a certain sympathy with his round statement: 'Till I see another round number, grounded upon many observations, nearer than 30, I hope to have done pretty well in multiplying our Burials by 30 to find the number of the People.'

With this I may conclude the analysis of Petty's statistical work. It will, I think, soon be clear enough that it is not of the calibre of Graunt's. Yet I cannot take leave of it without something of an *ave*. Careless, happy-go-lucky, tendentious; yes, all that. But anybody who has felt the exhilaration, to which Francis Galton owned, in the doing of sums concerning biological problems, feels his heart warmed by the arithmetical knight errant who had so many statistical adventures.

(*To be continued*)

FIDUCIAL ARGUMENT AND THE THEORY OF CONFIDENCE INTERVALS

By J. NEYMAN

University of California, Berkeley, California

CONTENTS

1. Introduction

The theory of confidence intervals was started by the present author about 1930. At that time it was taught in lectures given both at the University and at the Central College of Agriculture, Warsaw, Poland. The theory found immediate practical applications, and before any theoretical paper was published, a booklet (Pytkowski, 1932) appeared giving numerical confidence intervals for means and for regression coefficients. The term 'confidence interval' is a translation of the original Polish 'przedział ufności'. The author's theoretical results appeared two years later (Neyman, 1934). At almost the same time the first tables and graphs of confidence intervals were published (Clopper & Pearson, 1934) in a paper which gave a remarkably clear explanation of the difference between the new approach to the problem of estimation and the old one, by means of Bayes's theorem.

The first publication on fiducial argument (Fisher, 1930) anticipated the booklet of Pytkowski by two years. The present author overlooked this article for some time. However, when preparing his paper of 1934, he was already acquainted with it and also with the next paper (Fisher, 1933) on a similar subject. Although Fisher's method of approach was entirely different from the author's, the numerical identity of Fisher's fiducial limits with the confidence limits in the author's theory, and also some of Fisher's early comments, suggested to the author that the two theories are essentially the same. Accordingly, and owing to the difference in dates of publications, the author considered his own work as an

extension of the previous results of Fisher. This was clearly stated in the author's paper of 1934.

Apart from the above points of agreement the author had found certain passages and conceptions in the publications of Fisher which were difficult for him to understand and to reconcile with what was essential in the theory of confidence intervals. They included 'fiducial probability' and 'fiducial distribution of a parameter'. However, the author was inclined to think that these were, more or less, *lapsus linguae*, difficult to avoid in the early stages of a new theory. This attitude was clearly expressed in the paper of 1934. That paper was read before a meeting of the Royal Statistical Society and was followed by a public discussion recorded in the Society's *Journal*. Fisher took part in the discussion, and it was a great surprise to the author to find that, far from recognizing them as misunderstandings, he considered fiducial probability and fiducial distributions as absolutely essential parts of his theory. As a result, the author began to doubt whether the two theories were, in fact, equivalent. These doubts were only increased by Fisher's insistence that the calculation of fiducial distributions and fiducial limits must be limited to cases where sufficient statistics exist (Fisher, 1936), and by his warnings against inconsistencies in the theory of confidence intervals.

When questioned on the subject, the author could not conceal his doubts and they were published (Neyman, 1938*a*). Subsequent publications by other authors appear to be divided. Some, e.g. the very important papers by Wald (1939) and by Wald & Wolfowitz (1939), deal with the theory of confidence intervals, entirely ignoring fiducial theory. Others (Starkey, 1938; Sukhatme, 1938; Yates, 1939), at the other extreme, work on the ground of fiducial argument and ignore the confidence intervals. There is also an intermediate group of authors with an almost continuous spectrum of opinions. Pitman (1939), in a very interesting paper on estimation of location and scale parameters, states that the two theories 'are essentially the same and that their two points of view are both necessary for a full comprehension of the theory of estimation'. And a few pages further: 'I at first called it the fiducial probability function, but finally decided to shorten the name by dropping the word "probability".'

Next we find the statement (Bartlett, 1939) that 'by a distribution of fiducial type we shall mean a distribution providing at least confidence intervals in the sense of Neyman'. This statement is used in an argument (Bartlett, 1936, 1939) that, as a distribution deduced by Fisher (1936) does not seem to provide confidence limits, there must be some error in the deduction. A similar point of view, but with a stronger leaning towards confidence intervals, is expressed by Welch (1939). In this paper various general claims of Fisher are analysed, essentially from the point of view of confidence intervals, and tested on appropriate examples. Among other things it is found that the fears of inconsistencies in the theory of confidence intervals are unfounded.

A quite different school of thought is represented by Jeffreys (1940), according to which the fiducial approach to the problem of estimation is completely equivalent with that by inverse probability.

Fisher (1937, 1939*a*, 1939*b*) and Yates (1939) emphatically deny that there is an error in Fisher's paper of 1936. On the contrary, it is said that the results then published were obscured by the controversy arising from Bartlett's confusion about the nature of fiducial argument. Also, especially in earlier papers (1930, 1933, 1936), Fisher is equally emphatic on the distinction between the fiducial and the inverse probability approaches to the problem of estimation.

The above survey shows that there is an interesting divergence of opinions as to what is essential in the fiducial theory in general and as to whether it is in any way connected with the theory of confidence intervals. The perusal of all the literature quoted does not allow the present author to form any precise opinion as to the first of these questions. On the other hand, there now seems to be sufficient ground for answering the second, concerning the relationship between the two theories. The purpose of the present paper is to show that there is none. The relevant points concerning this question, which were possible to establish on the ground of earlier literature, are explained in excellent papers by Pearson (1939) and Welch (1939), with the final conclusion that, in spite of various differences, the two theories are closely related. However, fresh evidence provided by papers of Fisher (1939*a*, 1939*b*) and Yates (1939) shows that no such relation exists and that the authors suspecting it were misled by the incompleteness of earlier writings concerning fiducial argument.

As a result of the present paper it may be found expedient, for the sake of clarity, to avoid confusion of terminologies appropriate to the two theories. Instead of writing, as some authors do, on 'fiducial *or* confidence' limits, it may be preferable to discuss 'fiducial limits' or 'confidence limits', as the case may be, separately.

2. Basic ideas in the theory of confidence intervals

The key to understanding the theory of confidence intervals is in being clear about what might be called the classical point of view in the theory of probability. This theory was originally built up to answer questions about *how frequently* a given combination of throws will occur in a long series of games of dice. Thus, the probability of a certain combination found to be, say, 1/5, implies that this combination would appear in about 20 % of a long series of actual games. This agreement may, but need not, be observed. In the latter case, we would say that the assumptions underlying the deduction were not realized by the actual experiments. The dice used were perhaps 'biased', and so forth. The point is that, whenever it is said that a given set of probabilities does refer to some phenomena, then it is understood that the relative frequencies of various aspects of the phenomena,

in a long series of trials, are approximately equal to corresponding probabilities. This is just what the author calls the classical point of view in the theory of probability. It is excellently explained by v. Mises (1939), but is more general than the definition of probability adopted by that author.*

Apart from the classical point of view on probability, there is another. It considers the probabilities as measures of rational belief in the truth of a given proposition. Here the agreement between the probability and some relative frequency is not essential.

The theory of confidence intervals was built up to give a solution of problems of estimation which would have a clear frequency interpretation, characteristic of the classical point of view. Consider a set E of n observable random variables, $x_1, \ldots, x_n$, and assume as given that the function $p(E \mid \theta_1, \theta_2, \ldots, \theta_s)$ represents its elementary probability law. Here $\theta_1, \ldots, \theta_s$ represent certain parameters whose values are unknown.

The above should be interpreted as follows. There are some actual trials T which are able to determine the values of the x's. There are also some numbers $\vartheta_1, \vartheta_2, \ldots, \vartheta_s$, unknown to us, such that, whatever be a region w in the space of the x's, the integral of $p(E \mid \vartheta_1, \vartheta_2, \ldots, \vartheta_s)$ taken over this region is approximately equal to the relative frequency with which the point E, as determined by the trials T, falls within that region w. The problem of estimating one of the parameters, e.g. θ_1, consists in using just one system of the x's as determined by the trials T to calculate ϑ_1 approximately. Alternatively, it may consist in calculating an interval $(a, a+d)$ which 'presumably' covers ϑ_1.

The original approach to this problem is based on Bayes's theorem. Denote by $p(\theta_1, \theta_2, \ldots, \theta_s)$ the elementary probability law of the θ's. Then

$$p(\theta_1, \theta_2, \ldots, \theta_s \mid E') = \frac{p(\theta_1, \ldots, \theta_s)\, p(E' \mid \theta_1, \ldots, \theta_s)}{\int \cdots \int p(\theta_1, \ldots, \theta_s)\, p(E' \mid \theta_1, \ldots, \theta_s)\, d\theta_1, \ldots, d\theta_s} \qquad (1)$$

will be the relative probability law, or the probability law *a posteriori* of all the θ's given the observed system E' of the values of the x's. It can be used to calculate the most probable value of θ_1. Alternatively, given a number $d > 0$, the law can be used to find the interval $(a, a+d)$ such that the *a posteriori* probability

$$P\{a+d > \theta_1 > a \mid E'\}$$

is greatest.

Our attitude towards this kind of solution, dictated by the classical point of view on probability, depends on circumstances and may be twofold.

The circumstances of the problem may imply not only that the x's but also that the θ's are random variables and that the function $p(\theta_1, \ldots, \theta_s)$ could be used to

* It will be noticed that the classical point of view or probability does not imply any particular definition of that concept. It is not suggested that the ono adopted by v. Mises is the only one that could be consistently used.

calculate the relative frequencies of various combinations of values of the θ's. Such situations are rare, but they do occasionally occur, especially in problems of genetics and of mass production. If the function $p(\theta_1, \ldots, \theta_s)$ is implied by the problem considered, then the probability $P\{a+d>\theta_1>a \mid E'\}$ has a clear frequency interpretation, as follows. Imagine a long sequence, S, of cases where the θ's vary according to the above law and the x's are determined by the particular trials considered. Pick from this sequence S a subsequence $S(E')$ of such trials in which the experiments determined the same system of values of the x's, namely, the system E'. Naturally, the value of θ_1 in cases belonging to $S(E')$ would vary. But, if the functions $p(E \mid \theta_1, \ldots, \theta_s)$ and $p(\theta_1, \ldots, \theta_s)$ do have the presumed relation to the trials considered, it will be found that among all the intervals of length d, the interval $(a, a+d)$ will contain the value of θ_1 more frequently than any other, and that this frequency will be approximately equal to $P\{a+d>\theta_1>a \mid E'\}$. It follows that, if the function $p(\theta_1, \ldots, \theta_s)$ is implied by the circumstances of the problem of estimation, the use of the formula (1) is perfectly legitimate from the point of view of the classical theory of probability.

The situation is quite different when the circumstances of the problem do not imply the *a priori* probability law. This is most frequently the case. Moreover, usually there are serious difficulties in considering the θ's as random variables. Jeffreys (1939) advises the use of formula (1) also in such cases, with a function $p(\theta_1, \ldots, \theta_s)$ invented for the purpose. He claims that the conclusions drawn in this way are valid, provided that the function used is just the one that he suggests. The present author would not question this statement on condition that the word 'valid', or any other such description, is not given any significance beyond that described above. In other words, there seems to be no reason why we should not agree to *call* the above conclusions 'valid in the sense of Jeffreys'. On the other hand, it seems essential to be clear that any probability calculated from (1), with any function $p(\theta_1, \ldots, \theta_s)$ not implied by the actual problem, need not and, generally, will not have any relation to relative frequencies. It will not be the probability in the classical sense of the word and, therefore, persons who would like to deal only with classical probabilities, having their counterparts in the really observable frequencies, are forced to look for a solution of the problem of estimation other than by means of the theorem of Bayes.

This solution (Neyman, 1937, 1938*b*) may be obtained as follows. Consider the case where the circumstances imply that the x's, forming a system E, are random variables with the probability law $p(E \mid \theta_1, \theta_2, \ldots, \theta_s)$, where $\theta_1, \theta_2, \ldots, \theta_s$ are unknown. Denote by $\underline{\theta}(E)$ and $\overline{\theta}(E)$ two functions of the x's. Obviously, if E is random then these functions will also be random variables.

Definition 1. *If the functions $\underline{\theta}(E)$ and $\overline{\theta}(E)$ possess the property that, whatever be the possible value ϑ_1 of θ_1 and whatever be the values of the unknown parameters $\theta_2, \theta_3, \ldots, \theta_s$, the probability*

$$P\{\underline{\theta}(E) \leqslant \vartheta_1 \leqslant \overline{\theta}(E) \mid \vartheta_1, \theta_2, \ldots, \theta_s\} \equiv \alpha, \qquad (2)$$

then we will say that the functions $\underline{\theta}(E)$ *and* $\overline{\theta}(E)$ *are the lower and the upper confidence limits of* θ_1, *corresponding to the confidence coefficient* α. *The interval* $(\underline{\theta}(E), \overline{\theta}(E))$ *is called the confidence interval for* θ_1.

In spite of the complete simplicity of the above definition, certain persons have difficulties in following it. These difficulties seem to be due to what Karl Pearson (1938) used to call routine of thought. In the present case the routine was established by a century and a half of continuous work with Bayes's theorem. It may be useful, therefore, to give a few illustrations.

Assume that $s = 2$, that θ_1 may have only the five values 1, 2, 3, 4, and 5, and that, at the same time, θ_2 may vary continuously between zero and 1. To satisfy Definition 1, the only requirement on the functions $\underline{\theta}(E)$ and $\overline{\theta}(E)$ is that

$$P\{\underline{\theta}(E) \leqslant \vartheta \leqslant \overline{\theta}(E) \mid \vartheta, \theta_2\} \equiv \alpha \tag{3}$$

for all values of $\vartheta = 1, 2, 3, 4$, and 5, and for θ_2 varying between (0, 1). The probabilities (2) and (3) are, therefore, *not* the probabilities of θ_1 falling within any limits. On the contrary, they are the probabilities of the functions $\underline{\theta}(E)$ and $\overline{\theta}(E)$ falling on both sides of a specified number ϑ. These probabilities are to be calculated from the given function $p(E \mid \theta_1, \theta_2)$ with the value of θ_1 set equal to the same number ϑ. The result must be totally independent of the values of $\theta_2, \ldots, \theta_s$ and must equal α.

It is known (Neyman, 1935*b*; Feller, 1938) that in certain cases no such functions $\underline{\theta}(E)$ and $\overline{\theta}(E)$ exist. Then there are ways of modifying the formulation of the problem, for example, requiring that the probability on the left of (2) be at least equal to α, and so forth. In other cases, there will be an infinity of pairs of confidence limits all corresponding to the same α. In this case, the practical statistician is at liberty to choose among them.

Let us now consider the frequency interpretation of the solution of the problem of estimation by means of confidence intervals. Suppose that some two functions $\underline{\theta}(E) \leqslant \overline{\theta}(E)$ possess property (2) with some large value of α, say $\alpha = 0{\cdot}99$. Their use in practice would consist of (i) observing the value E' of the x's, (ii) calculating the corresponding values of the confidence limits $\underline{\theta}(E')$ and $\overline{\theta}(E')$, and (iii) *stating* that the true value ϑ_1 of θ_1 lies between $\underline{\theta}(E')$ and $\overline{\theta}(E')$. The justification is simple and perfectly in line with the classical point of view of probability: in many applications, the relative frequency of cases in which the statement $\underline{\theta}(E) \leqslant \vartheta_1 \leqslant \overline{\theta}(E)$ is correct will be approximately equal to $\alpha = 0{\cdot}99$, whether or not the parameters for estimation are the same in all cases.

The word 'stating' above is put in italics to emphasize that it is not suggested that we can 'conclude' that $\underline{\theta}(E') \leqslant \vartheta_1 \leqslant \overline{\theta}(E')$, nor that we should 'believe' that ϑ_1 is actually between $\underline{\theta}(E)$ and $\overline{\theta}(E)$. In the author's opinion, the word 'conclude' has been wrongly used in that part of statistical literature dealing with what has been termed 'inductive reasoning'. Moreover, the expression 'inductive reasoning' itself seems to involve a contradictory adjective. The word 'reasoning' generally

seems to denote the mental process leading to knowledge. As such, it can only be deductive. Therefore, the description 'inductive' seems to exclude both the 'reasoning' and also its final step, the 'conclusion'. If we wish to use the word 'inductive' to describe the results of statistical inquiries, then we should apply it to 'behaviour' and not to 'reasoning'. The fact that a given pair of functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ satisfies the identity (2) may be 'deduced' from the properties of the function $p(E \mid \theta_1, \ldots, \theta_s)$. Earlier trials may show characteristics in the empirical distribution of the x's which seem in agreement with the function $p(E \mid \theta_1, \ldots, \theta_s)$. On these grounds, after observing the values of the x's in a case where the θ's are unknown and calculating $\underline{\theta}(E')$ and $\bar{\theta}(E')$, we may *decide* to behave as if we actually knew that the true value ϑ_1 of θ_1 were between $\underline{\theta}(E')$ and $\bar{\theta}(E')$. This is done as a result of our *decision* and has nothing to do with 'reasoning' or 'conclusion'. The reasoning ended when the functions $\underline{\theta}(E)$ and $\bar{\theta}(E)$ were calculated. The above process is also devoid of any 'belief' concerning the value ϑ_1 of θ_1. Occasionally we do not behave in accordance with our beliefs. Such, for example, is the case when we take out an accident insurance policy while preparing for a vacation trip. In doing so, we surely act against our firm belief that there will be no accident; otherwise, we would probably stay at home. This is an example of inductive behaviour.

Obviously, if there are many different pairs of functions, $\underline{\theta}(E)$ and $\bar{\theta}(E)$, all corresponding to the same α, our choice of the one to use must be based on the detailed study of their properties. For example, if it appears that the difference between one pair, $\bar{\theta}_1(E) - \underline{\theta}_1(E)$, is always (or most frequently) smaller than that between some other pair, then we would probably prefer to use the first. The problem of determining the confidence limits and of studying their properties forms the subject of the theory of confidence intervals.

3. Necessary and sufficient conditions for a pair of functions to be confidence limits

Let $a(E) \leqslant b(E)$ be any two single-valued functions of the x's determined for all possible systems of their values. Denote by W the space of the x's and by ϑ_1 one of the possible values of θ_1. Finally, let $A(\vartheta_1)$ denote the region in the space W composed of all points E which satisfy the double inequality,

$$a(E) \leqslant \vartheta_1 \leqslant b(E). \tag{4}$$

It was proved (Neyman, 1937) that for the two functions, $a(E)$ and $b(E)$, to be the lower and upper confidence limits for the parameter θ_1, it is necessary and sufficient that, whatever be the possible value ϑ_1 of θ_1, the probability

$$P\{E \epsilon A(\vartheta_1) \mid \theta_1 = \vartheta_1\} \equiv \alpha. \tag{5}$$

The identity refers to the arbitrary variation of $\theta_2, \ldots, \theta_s$.

This condition will be used below to show that a certain pair of functions does not represent the confidence limits. For this purpose, the following steps will be taken: We shall select a convenient value ϑ_1 of the estimated parameter θ_1 and

determine the region $A(\vartheta_1)$ as in (4). Next, we shall substitute this same value ϑ_1 instead of the parameter θ_1 in the elementary probability law of the variables considered, getting $p(E \mid \vartheta_1, \ldots, \theta_s)$. This last function will be integrated over $A(\vartheta_1)$ to find the probability $P\{E \epsilon A \mid \theta_1 = \vartheta_1\}$ as in the left-hand side of (5). But this integral will be dependent on the values of the other parameters involved, showing that the identity (5) is not satisfied. The conclusion will be that the particular functions considered are not confidence limits.

4. Differences between the theory of confidence intervals and the theory of fiducial argument

In this section we will consider examples treated both from the point of view of confidence intervals and of fiducial argument. These will be selected to illustrate both the conceptual and the numerical differences between the two theories.

(i) *Evidence of conceptual differences between the two theories.* The first results obtained concerning confidence intervals (Neyman, 1934) refer to the case where all the n observable variables x_i are mutually independent, normally distributed, have the same though unknown standard error σ, and expectations $\mathscr{E}(x_i)$ which are linearly connected with some $s < n$ unknown parameters $p_1, p_2, \ldots, p_s$, so that

$$\mathscr{E}(x_i) = a_{i1}p_1 + a_{i2}p_2 + \ldots + a_{is}p_s. \tag{6}$$

Here the a's are supposed to be known and to form a non-singular matrix. Denote by θ any linear combination of the same p's, that is

$$\theta = b_1p_1 + b_2p_2 + \ldots + b_sp_s, \tag{7}$$

with known b's not all equal to zero. In these circumstances, a confidence interval for θ is given by

$$F - St_\alpha \leqslant \theta \leqslant F + St_\alpha, \tag{8}$$

where F denotes the best unbiased estimate of θ (David & Neyman, 1938), S the estimate of the standard error of F, and t_α the value of the 'Student'-Fisher t corresponding to the number of degrees of freedom $n-s$ and to $P = 1-\alpha$. The application of more recent theory (Neyman, 1935*b*) shows that the confidence intervals (8) have distinct advantages over any others by satisfying the definition (Neyman, 1937) of the 'short unbiased system of type B_1'. Without entering into these details, we shall consider the particular case where $s = 1$, $a_{i1} = 1$ and $b_1 = 1$. This will be the case if all the x's come from the same unknown normal population and it is desired to estimate its mean, $\theta = \mathscr{E}(x_i)$. In that case $F = \bar{x}$ and

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n(n-1)}. \tag{9}$$

As mentioned, the general confidence interval (8) was discussed in lectures about 1930, and in 1932 a publication appeared using the concept and the formula (8).

As far as is known, the first full discussion of the corresponding result in the fiducial theory was given by Fisher a few years later (Fisher, 1935, 1936), and here is the relevant passage from the second paper.

If a sample of n observations, $x_1, \ldots, x_n$, has been drawn from a normal population having a mean value μ, and if from the sample we calculate the two statistics $\bar{x} = \Sigma x_i/n$ and $s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$, ..., 'Student' has shown (1925)* that the quantity t, defined by the equation

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s}, \tag{10}$$

is distributed in different samples in a distribution dependent only from the size of the sample, n. It is possible, therefore, to calculate, for each value of n, what value of t will be exceeded with any assigned frequency, P, such as 1 % or 5 %. These values of t are, in fact, available in existing tables (Fisher, 1925–34).

It must now be noticed that t is a continuous function of the unknown parameter, the mean, together with observable values, $\bar{x}$, s and n, only. Consequently the inequality $t > t_1$ is equivalent to the inequality

$$\mu < \bar{x} - st_1/\sqrt{n}, \tag{11}$$

so that this last inequality must be satisfied with the same probability as the first. This probability is known for all values of t_1, and decreases continuously as t_1 is increased. Since, therefore, the right-hand side of the inequality takes, by varying t_1, all real values, we may state the probability that μ is less than any assigned value, or the probability that it lies between any assigned values, or, in short, its probability distribution, in the light of the sample observed.

It is of some importance to distinguish such probability statements about the value of μ, from those that would be derived by the method of inverse probability, from any postulated knowledge of the distribution of μ in the different populations which might have been sampled.... To distinguish it from any of the inverse probability distributions derivable from the same data it has been termed the *fiducial* probability distribution, and the probability statements which it embraces are termed statements of fiducial probability.

In the next section we shall analyse the above passage in detail and show exactly where and how it conflicts with the classical theory of probability and thus with the theory of confidence intervals. Here we will mention only that it is ambiguous. Just this kind of ambiguity, which is also found in the earlier papers (Fisher, 1930, 1933), is probably responsible for a number of authors, including the present one, thinking that the fiducial theory and the theory of confidence intervals are linked.

In a few years it was found necessary to reinterpret formula (11). This was done by Fisher himself (1939 *b*) and, somewhat more clearly but on the same lines, by Yates (1939). It will be seen from the following quotation from Yates's paper that the above passage by Fisher certainly does not contain everything which is *now* considered essential in the fiducial theory and that the presumption of any link between the latter and the theory of confidence intervals is unfounded. Yates's more relevant sentences are italicized by the present author.

While explaining the meaning of the fiducial distribution of the mean μ of a normal population, Yates mentions that the fiducial distribution of σ^2 is given by

$$\frac{1}{\sigma^2} = \frac{\chi_0^2}{\Sigma(x_i - \bar{x})^2}, \tag{12}$$

where χ^2 has its usual distribution with $n-1$ degrees of freedom.

* Actually, of course, this result appeared earlier ('Student', 1908).

It can then be shown that, for a value of μ equal to μ_ε and a given s, the value of $\bar{x}$ in subsequent samples would be as small as that observed in a fraction ϵ of the samples, *provided that the actual distribution of σ^2 is the same as the fiducial distribution given above.*

In this form, however, the statement is open to objection on the ground that in subsequent samples σ may in fact be distributed in any manner, and that s will certainly vary from sample to sample. To avoid this objection we must frankly recognize *that we have here introduced a new concept into our methods of inductive inference, which cannot be deduced by the rules of logic from already accepted methods*....That is...the form of fiducial statement which is implicit in the t test as ordinarily used by practical experimenters....It must be recognized as essentially different from the statement that t will exceed t_ε in a fraction ϵ of all experiments. The latter is true for any given fixed σ or any set of σ's. The former (i.e. the fiducial statement, J.N.) *is true for a given s when σ is taken to be fiducially distributed in the appropriate distribution*....The logical difference between the two approaches (fiducial and inverse probability, J.N.) should, however, be recognized. The approach by inverse probability enables fiducial statements about μ to be derived from the classical theory of probability, without the introduction of any new principle, but only at the cost of postulating a particular *a priori* distribution of σ. *In the fiducial approach such a priori postulation is regarded as inadmissible, but in order to discard it a new principle, that of utilizing the fiducial distribution of σ, must be introduced*....Once the principle is accepted it is possible, given $\bar{x}$ and s, to make formal and exact statements of the fiducial type about μ which are independent of all prior knowledge of σ. *If the principle is not accepted, then it appears that we must either assume an a priori distribution of σ, or deny that there is any possibility of making fiducial statements about μ.*

The present author is unable to understand the exact meaning of what is called 'fiducial statements about μ'. However, his conclusion is that their conceptual nature must be quite different from that dealt with in the theory of confidence intervals. This conclusion is based on the fact that all the difficulties described by Yates as inherent in the fiducial theory are non-existent in the theory of confidence intervals. Applications of the latter require no new principle 'which cannot be deduced by the rules of logic', no assumption that this or that unknown parameter follows any specified distribution, and have no connexion with Bayes's theorem. To make the situation absolutely clear, imagine a sequence of normal populations $\pi_1, \pi_2, \ldots, \pi_m, \ldots$, with their means $\theta_1, \theta_2, \ldots, \theta_m, \ldots$ and their standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_m, \ldots$. Imagine that out of each population π_m we have a random sample Σ_m of n individuals, with its mean $\bar{x}_m$ and an estimate of the corresponding variance S_m^2 as in (9). The theory of confidence intervals guarantees that the relative frequency with which $\bar{x}_m - t_\alpha S_m$ will fall short of the corresponding θ_m and, at the same time $\bar{x}_m + t_\alpha S_m$ will exceed this same number θ_m, will be, within an error of sampling, equal to α. An incredulous reader may easily check this by a sampling experiment. In this he will be at liberty to keep θ_m and/or σ_m constant, or to vary them at his pleasure, without any restriction. Of course, the distributions of the populations sampled should be more or less normal and the sampling should be random. It follows from the above passages of Yates that if the requirements above are satisfied and no new principles accepted, then we have to deny that there is any possibility of making fiducial statements about θ_m. If so, then the nature of the latter is different from those involved in the application of the theory of confidence intervals.

The comparison of the above comments by Yates with those of Fisher gives a curious impression. Where Yates sees so many difficulties and restrictions, Fisher mentions none. Yet this very publication of Yates is fully endorsed by Fisher (1939 *b*).

(ii) *Numerical differences between the two theories.* Besides establishing the existence of conceptual differences, it is essential to show that the two theories may give different numerical results. We may conclude from the discussion above that the application of confidence intervals requires fewer restrictions. But there is a logical possibility that, when both theories are applicable, they give the same numerical result. The following example shows that this is not the case and that fiducial limits need not satisfy the definition of confidence limits.

The example that we are going to discuss refers to the problem of estimating the difference, say δ, between the means of two populations of which it is known only that both are normal. Denote by

$$\left.\begin{matrix} x_{1,1}, & x_{1,2}, \ldots, & x_{1,n}, \\ x_{2,1}, & x_{2,2}, \ldots, & x_{2,n'}, \end{matrix}\right\} \tag{13}$$

two random samples to be drawn from these populations and let $n \leqslant n'$. The confidence limits for δ have been very elegantly obtained by Bartlett. He did not publish his results himself but they are briefly mentioned in a paper by Welch (1938). The tendency towards a greater generality of presentation resulted in certain complications. The following is a less general but simplified statement of the results.* Assume that the x's in (13) are numbered in the order in which they will be given by observation. Otherwise, randomize the second series. Next calculate n differences

$$u_i = x_{1,i} - x_{2,i} \quad (i = 1, 2, \ldots, n). \tag{14}$$

If $\mathscr{E}(x_{1,i}) = \theta + \delta$ and $\mathscr{E}(x_{2,i}) = \theta$, then $\mathscr{E}(u_i) = \delta$. If the S.D.'s of the two populations sampled are σ and σ', then the S.E. of u_i will be $(\sigma^2 + \sigma'^2)^{\frac{1}{2}}$. The consecutive u's will be normal and independent and the problem of estimating the difference between the means of two normal populations will be reduced to that of estimating the mean of one population of the u's. Its solution is given by the confidence interval

$$\bar{u} - St_{(\alpha)} \leqslant \delta \leqslant \bar{u} + St_{(\alpha)}, \tag{15}$$

where S has an obvious meaning and $t_{(\alpha)}$ is to be taken with $n-1$ degrees of freedom.

Again, an experiment consisting in repeated sampling of pairs of normal populations will show that, whatever be θ, δ, σ, σ', whether constant or varying in an absolutely arbitrary manner, the relative frequency of cases in which the statement about δ in the form of (15) will be true will be approximately equal to α. The above solution of the problem, elegant as it is, is only a partial one. The results

* Apart from these, the same author has obtained certain relevant results referring to the case where $n = n' = 2$ (Bartlett, 1936).

of Bartlett do not tell us whether the family of systems of confidence intervals found by him exhausts all the possibilities and whether it is possible to construct intervals which would be, in one sense or another, shorter than those given by (15). These are interesting and important problems and we may hope to have them solved.

A result in fiducial theory corresponding to, but not equivalent with, formula (15) has been published by Fisher (1936):

Let us suppose that a sample of n observations has yielded a mean, $\bar{x}$, and an estimated variance of the mean, s^2, so that $s^2 = \Sigma(x_i - \bar{x})^2/n(n-1)$; then we know that if μ is the mean of the population

$$\mu = \bar{x} + st, \tag{16}$$

where t is distributed in 'Student's' distribution. Similarly, for the mean of a second population, of which we have n' observations, we may write

$$\mu' = \bar{x}' + s't', \tag{17}$$

where t' is distributed in 'Student's' distribution with $n'-1$ degrees of freedom, independently of t. If now

$$\mu' - \mu = \delta, \quad \bar{x}' - \bar{x} = d, \tag{18}$$

we find that

$$\epsilon = \delta - d = s't' - st, \tag{19}$$

and since s' and s are known, the quantity represented on the right has a known distribution, though not one which has been fully tabulated. The equation may be written

$$\epsilon = \sqrt{(s^2 + s'^2)}\,(t' \cos R - t \sin R), \tag{20}$$

where $\tan R = s/s'$, so that R is a known angle. If t and t' be taken as the co-ordinates of a point on a plane, the frequency of the observations falling within any area of tho plane is calculable. The points for which ϵ has any given value lie on a straight line, at a distance from the origin $\pm\epsilon/(s^2+s'^2)^{\frac{1}{2}}$, and making an angle R with the axis of t. The fiducial probability that ϵ exceeds any given value is the frequency in the area above this line. If n and n' are both increased, the distribution of ϵ tends to be normal and independent of R; when R is 0° or 90° the distribution is of 'Student's' form. In general it involves n, n', and R and for any chosen probability, therefore, requires a table of triple entry.

As the reader will notice, no restrictions are mentioned and it is not suggested that for the practical application of the results any assumption is needed concerning the variability of the variances of the populations sampled. Neither is there any suggestion of any new principle that may be involved. We will return to this point below.

Following the publication of Fisher just quoted, and on his advice, Sukhatme published a table (Sukhatme, 1938). The quantity tabled may be denoted by $f(n, n', R)$ and represents the root of the equation

$$\int_{-\infty}^{+\infty}\left\{G(t)\int_{\kappa}^{+\infty} H(t')\,dt'\right\}dt = 0{\cdot}025, \tag{21}$$

where $G(t)$ and $H(t')$ are 'Student's' distributions with $n-1$ and $n'-1$ degrees of freedom respectively, while

$$\kappa = \frac{f(n, n', R)}{(s^2+s'^2)^{\frac{1}{2}} \cos R} + t \tan R. \tag{22}$$

It follows from the context that $f(n, n', R)$ so calculated is the value such that the fiducial probability of its being exceeded by $|\epsilon|/(s^2+s'^2)^{\frac{1}{2}}$ is equal to 0·05. In other words, the values $f(n, n', R)$ are the fiducial 5 % limits of $|\epsilon|/(s^2+s'^2)^{\frac{1}{2}}$. As $\epsilon = \delta - d$, if the presumption that the fiducial limits necessarily lead to confidence intervals be true then this means that the double inequality

$$\bar{x}' - \bar{x} - f(n, n', R)\sqrt{(s^2+s'^2)} \leqslant \delta \leqslant \bar{x}' - \bar{x} + f(n, n', R)\sqrt{(s^2+s'^2)} \tag{23}$$

must be the confidence intervals for $\delta = \mu' - \mu$. But it is easy to see that the functions on the extreme parts of (23) do not satisfy the conditions, explained in § 3 above, necessary and sufficient for them to be the confidence limits. Take $\delta = 0$ and denote simply by A the region in the space of the x's including all the points in which the inequality (23) is satisfied. Take the probability law of the x's and put $\delta = 0$ in it, that is, $\mu' = \mu$. It will be seen that the integral $I(A)$ of this probability law taken over A depends on the ratio $\rho = \sigma/\sigma'$ of the two σ's appropriate to the two populations sampled and, thus, that it does not satisfy the identity (5).

Condition (23) defining the region A does not involve the particular x's but only the means $\bar{x}$, $\bar{x}'$, and the variances s^2 and s'^2. Consequently, to calculate $I(A)$ we may start with the probability law of those four variables

$$p(\bar{x}, \bar{x}', s, s') = \frac{c}{\sigma^n \sigma'^{n'}} s^{n-2} s'^{n'-2}$$
$$\times \exp\left[-\frac{n(\bar{x}-\mu)^2}{2\sigma^2} - \frac{n'(\bar{x}'-\mu)^2}{2\sigma'^2} - \frac{n(n-1)s^2}{2\sigma^2} - \frac{n'(n'-1)s'^2}{2\sigma'^2}\right], \tag{24}$$

where c is a purely numerical constant and does not involve any of the parameters. This function must be integrated over the region A defined by (23) or by the equivalent inequality

$$\frac{|\bar{x}' - \bar{x}|}{\sqrt{(s^2+s'^2)}} \leqslant f(n, n', R). \tag{25}$$

In dealing with it, we have to remember that R is not a constant but is connected with s and s' by the equation $\tan R = s/s'$. The required integral, or probability, of $\bar{x}$, $\bar{x}'$, s, and s' satisfying (25) will be more easily calculated if we introduce a new system of variables, u, v, R, and s_0. These will be connected to the old system as follows:

$$\left.\begin{aligned} \bar{x} &= \mu + us_0 \sin R, \\ \bar{x}' &= \mu + vs_0 \cos R, \\ s &= s_0 \sin R, \\ s' &= s_0 \cos R. \end{aligned}\right\} \tag{26}$$

The Jacobian J of the transformation is easily found to be

$$J = s_0^3 \sin R \cos R. \tag{27}$$

The limits of variation of the new variables are as follows:

$$\left.\begin{aligned} -\infty < u, v < +\infty, \\ 0 \leqslant s_0, \\ 0 \leqslant R \leqslant \tfrac{1}{2}\pi. \end{aligned}\right\} \tag{28}$$

The probability law of the new variables will be

$$p(u, v, s_0, R) = \frac{c}{\sigma^n \sigma'^{n'}} s_0^{n+n'-1} e^{-\frac{1}{2}(\psi^2 s_0^2)} \sin^{n-1} R \cos^{n'-1} R, \tag{29}$$

with
$$\psi^2 = \frac{nu^2 \sin^2 R}{\sigma^2} + \frac{n'v^2 \cos^2 R}{\sigma'^2} + \frac{n(n-1)\sin^2 R}{\sigma^2} + \frac{n'(n'-1)\cos^2 R}{\sigma'^2}. \tag{30}$$

Inequality (25) will be equivalent to

$$|v \cos R - u \sin R| \leqslant f(n, n', R). \tag{31}$$

As this does not involve s_0 the integration with respect to this variable can be carried out within the extreme limits of its variation. As a result further integrations may be performed on the probability law of u, v, R,

$$\begin{aligned} p(u, v, R) &= \int_0^\infty p(u, v, s_0, R)\, ds_0 \\ &= \frac{c}{\sigma^n \sigma'^{n'}} \frac{\sin^{n-1} R \cos^{n'-1} R}{\psi^{n+n'}}, \end{aligned} \tag{32}$$

where c is again a numerical constant.

Further integration may be conveniently carried out as follows. Substitute a new variable z for the variable v so that

$$v = \frac{z + u \sin R}{\cos R}, \quad \frac{\partial v}{\partial z} = \frac{1}{\cos R}. \tag{33}$$

Keep z constant within the limits $|z| \leqslant f(n, n', R)$ prescribed by (31) and integrate for u from $-\infty$ to $+\infty$. The result is

$$\begin{aligned} p(z, R) = {} & \frac{c \sin^{n-2} R \cos^{n'-2} R}{\sigma^{n-1} \sigma'^{n'-1} \sqrt{(n\sigma'^2 + n'\sigma^2)}} \\ & \times \left\{ \frac{nn'}{n\sigma'^2 + n'\sigma^2} z^2 + \frac{n(n-1)}{\sigma^2} \sin^2 R + \frac{n'(n'-1)}{\sigma'^2} \cos^2 R \right\}^{-\frac{1}{2}(n+n'-1)}. \end{aligned} \tag{34}$$

The integration is completed by an easy substitution for z

$$\begin{aligned} I(A) = {} & c\rho^{n'-1} \\ & \times \int_0^{\frac{1}{2}\pi} \left\{ \frac{\sin^{n-2} R \cos^{n'-2} R}{(n(n-1)\sin^2 R + n'(n'-1)\rho^2 \cos^2 R)^{\frac{1}{2}(n+n'-2)}} \int_0^{w f} \frac{dz}{(1+z^2)^{\frac{1}{2}(n+n'-1)}} \right\} dR, \end{aligned} \tag{35}$$

with $f = f(n, n', R)$ and

$$w^2 = \frac{nn'}{n\sigma'^2 + n'\sigma^2} \Bigg/ \left(\frac{n(n-1)}{\sigma^2} \sin^2 R + \frac{n'(n'-1)}{\sigma'^2} \cos^2 R \right). \tag{36}$$

By inspecting (35) it is more or less evident that $I(A)$ must depend on the value of ρ. However, to avoid any doubt in this respect, it was thought useful to calculate $I(A)$ for a few values of ρ. This was done by Miss Elizabeth Scott of the Statistical Laboratory, University of California, and it is a pleasure to record the author's indebtedness to her. The calculations involved supplementing the tables of Sukhatme for a denser set of values of R. The calculated values of $I(A)$ are:

$n = 12, \quad n' = 6$

ρ	$I(A)$
0·1	0·966
1·0	0·960
10·0	0·934

Thus the functions representing the fiducial limits for δ do not satisfy the conditions necessary and sufficient for them to be the confidence limits of the parameter in question. It follows that if pairs of normal populations forming a long sequence are sampled and the extreme parts of the double inequality (23) calculated, then the relative frequency of cases where the prediction of the value of δ by means of these inequalities will be correct need not be equal to the expected 0·95. It will depend on the value of ρ and, if this is uncertain, this frequency will be unknown. Subsequent comments by Fisher (Fisher, 1939*a*) seem to indicate that the frequency in question is expected to approach 0·95 only if the ratio ρ is not constant but follows a certain fiducial distribution. It is noteworthy that no such restriction is to be found in the original work quoted above. On the other hand, it is more or less in line with those restrictions formulated by Yates.

5. Views of M. S. Bartlett and R. A. Fisher

The controversy in which the main contributors are Bartlett (Bartlett, 1936, 1939) and Fisher (Fisher, 1937, 1939*a*, 1939*b*) seems to be based on a misunderstanding. Presuming that the fiducial limits are always equal to confidence limits, Bartlett was puzzled by Fisher's results concerning δ just quoted, and suspected an error. The subsequent elaborations by Fisher and Yates amount to a confirmation that the values of $f(n, n', R)$ as tabled by Sukhatme do not provide the confidence intervals. But both authors are emphatic that there is no error in the original deductions, and that Bartlett misunderstood the problem. It is unthinkable that these four unanimous papers are mistaken and, therefore, we must accept the conclusion that the presumption of intrinsic identity between fiducial and confidence limits is unfounded.

But it must be pointed out that, before the appeal to extra-logical principles

was published, there was much to be said in favour of the opinion that the solution of Fisher, as quoted above, and the work of Sukhatme both involved errors in the algebra of probability laws. It also seems that, apart from establishing that the fiducial theory and the theory of confidence intervals are distinct, it will be of some interest to analyse Fisher's work in detail and to point out exactly where and how it diverges from the rules of ordinary theory of probability on which the theory of confidence intervals is based.

When a system of observable phenomena is treated mathematically, it is essential to be clear on exactly what is assumed as given or as known. For example, when trying to calculate the area of land from a certain set of measurements, it is essential to be clear as to assumptions made concerning the shape of the land considered. The available data may be consistent with a number of such assumptions, e.g. that the surface considered is a plane or that it is spherical with a given radius, etc. Whichever of these hypotheses is accepted as given, the applications of the appropriate formulae will give mutually consistent results. But they would not generally be consistent if one part of the calculations were made on one hypothesis and another on a contradictory one. The differences may be small, but in mathematics there are really no 'small' nor 'large' inconsistencies. There are simply inconsistencies. Needless to say, the choice of exactly what is to be accepted as given must be made to attain the greatest conformity with empirical facts. But this is a question which need not be discussed here.

The above general principle also applies to the applications of probability. There we must be clear as to exactly what are the phenomena or the variables which we agree to consider as *random* in a given inquiry. In practice, of course, the random variable will be the one whose value at the moment is uncertain and is being determined 'by chance'. If X is considered as a random variable, the premises of the mathematical problem must include some assumptions as to the relative frequencies with which X assumes its possible values. These assumptions may vary in specificity, but they must be present in the premises.

Any number or variable which is not random must be clearly recognized as such. For some time such non-random numbers were called constants. This was more or less satisfactory with constant *numbers*. But Fréchet (Fréchet, 1937) has noticed that we may also consider *variables* which are not random and has invented useful terms to describe them. These are 'nombre certain', 'fonction certaine', etc. We will translate these terms by 'sure number' and 'sure function'. The thousandth digit in the expansion $\pi = 3{\cdot}1415\ldots$ is a sure number, although totally unknown to me. Denote by $f(n)$ the relative frequency of 0's among the first n digits of the same expansion of π. This will be a sure function. On the other hand, if $\phi(n)$ denotes the number of errors that may be made when calculating π to n places of decimals, then $\phi(n)$ *may be* considered as a random function of n. Considerations of this kind would imply those of a considerable sequence S of similar attempts to calculate π, by the same person or by different persons of a

specified category, in which the values of $\phi(n)$ will vary, as we shall say, at random. It is with respect to just such a sequence of determinations of the values of the function $\phi(n)$ that our probability statements will refer. For example, if we either start or finish our calculations with the probability equal to 0·25 of $\phi(n)$ being between any two sure numbers a, b, then the applicational statement is that about 25 % of the numbers of the sequence S satisfy the inequality $a < \phi(n) < b$.

It is important to notice that the sequence S *may* consist of just one member; then all the proportions relating this 'sequence' will have to be either 0 or 1. In other words, if the sequence of 'random' determinations consists of just one element, this element will have the property of a sure, not a random, object, in the usual sense of the word.

Now let us turn to the passage from Fisher's paper quoted above, p. 136, and try to see exactly what is supposed to be random there and what elements of the problem are treated as sure numbers or sure functions. These details in the set-up are not stated at the outset, but there is no difficulty in collecting them from appropriate passages in the paper. We first see that the function t of (10) is supposed to be 'distributed in different samples...'. This means that t is a random variable and that its randomness depends on what is found in those repeated samples, namely, the values of $\bar{x}$ and s. It follows that the probabilities concerning $\bar{x}$, s, and t refer to the sequence S of those 'different' samples. The sequence could not consist of just one sample because, in such a case, the 'distribution' of t would not be anything like 'Student's' law. The references to a normal population sampled and to 'Student's' law indicate, on the contrary, that the sequence S of samples is very large indeed, and that the distributions in it are comparable to those represented by continuous curves.

Up to this time we have not mentioned the population mean μ which is also involved in the expression of t. Obviously, this may be treated mathematically either as a random or as a sure number. Both methods of approach are at our disposal but, in order to avoid inconsistencies, we must be clear as to which one we follow. The indication of Fisher's choice is found a little further on in this article, in the place describing the distinction between the fiducial and the inverse probability approach: 'It is of some importance to distinguish such (fiducial) probability statement about the value of μ, from those that would be derived by the method of inverse probability from any postulated knowledge of the distribution of μ in the different populations which might have been sampled.' This sentence does not seem to leave any ground for doubt. In the fiducial approach we consider but one population sampled and no distribution of μ is postulated. Therefore, μ is a sure number and, if t is distributed according to 'Student's' law, it is a result of the appropriate variability of $\bar{x}$ and s alone.

The symbol t_1, which also comes into play, is obviously a sure variable capable of any real value between $-\infty$ and $+\infty$. We may select it as we wish and then obtain the probability $P(t_1)$ of the random variable t exceeding t_1 from tables.

Following the article, we will readily agree with Fisher that the inequality (11), namely, $\mu < \bar{x} - st_1/\sqrt{n}$, is equivalent to $t > t_1$ and that it must be satisfied with some probability $P(t_1)$. Now consider the phrase: 'Since, therefore, the right-hand side of the inequality (i.e. $\bar{x} - st_1/\sqrt{n}$) takes, by varying t_1, all real values, we may state the probability that μ is less than any assigned value, or the probability that it lies between any assigned values, or, in short, its probability distribution *in the light of the sample observed*.' From the point of view of ordinary logic and of ordinary theory of probability this phrase is inconsistent with the original set-up. The first inconsistency is involved in the words which are italicized, suggesting that $\bar{x}$ and s in the expression $\bar{x} - st_1/\sqrt{n}$ are not random but sure numbers, referring to one particular observed sample. As a matter of fact this same inconsistency appears earlier in the statement that $\bar{x} - st_1/\sqrt{n}$, by varying t_1, will run through all real numbers. If, as formerly, $\bar{x}$ and s are random with their variation appropriate to the sequence S, then, whatever value we choose to ascribe to t_1, say $t = 2$, the expression $\bar{x} - 2s/\sqrt{n}$ is also random and depends on the outcome of sampling.

Apart from this sudden shift in the meaning ascribed to $\bar{x}$ and s, there are two more inconsistencies. To see the first of them, let us follow Fisher, changing our minds about $\bar{x}$ and s and considering them as sure numbers, determined by one particular sample. In this case the inequality $\mu < \bar{x} - st_1/\sqrt{n}$ would contain no random elements at all: the first element, μ, is an unknown constant, the mean of a single population sampled, $\bar{x}$ and s are fixed by the sample observed, and t_1 is the value of the sure variable that we have chosen to consider. In these circumstances, the inequality may either be true or not true and the probability of its being true will equal unity or zero and have nothing to do with the probability or frequency $P(t_1)$ which this same inequality satisfies within a sequence S of many 'different' samples.

The last inconsistency refers, of course, to the point of view on μ. As we have seen above, it is first considered as a sure number, but the passage just quoted speaks of the probability of its lying between any assigned limits possible to determine from the values of $P(t)$. Assume $n = 4$ and that the sample observed gives $x = 10$ and $s = 2$. Select $t_1 = 0{\cdot}765$ and $t_1' = -0{\cdot}765$ so that $P(t_1) = 0{\cdot}25$ and $P(t_1') = 0{\cdot}75$. This would result in the supposed probability P' of μ lying between the limits $9{\cdot}235 \leqslant \mu < 10{\cdot}765$, being equal to 1/2. Trying to interpret this result in the light of the classical theory of probability, we have to conceive a sequence, say S', of cases in 50 % of which μ falls between the above limits. But exactly what could this sequence be? Either there is such a sequence and then we must also consider other populations 'which might have been sampled', *and postulate something about the distribution of μ**, or else the 'sequence' must be the degenerate one of one element only with the probability P' equal to either zero or unity, but never to 1/2.

These are the points previously mentioned by the author (Neyman, 1934),

* This is quite essential. Otherwise there would be an error in Bayes's theorem.

which, from the point of view of classical probability, represent conceptual inconsistencies. They are also present in the other passage of Fisher quoted on p. 139, but a similar analysis of that passage supplemented by what has subsequently been done by Sukhatme, will reveal errors in algebra of probability laws as well. These errors are particularly relevant from the point of view of the controversies between Bartlett and Fisher.

The quantities considered in this passage are all dependent on the population means μ and μ' and on the statistics $\bar{x}$ and s of one random sample and on $\bar{x}'$ and s' of the other. Our analysis will also require the consideration of the population variances σ^2 and σ'^2. We must start by deciding on the random or sure character of all these quantities. Fisher's remark that the two ratios

$$t = \frac{\mu - \bar{x}}{s} \quad \text{and} \quad t' = \frac{\mu' - \bar{x}'}{s'} \tag{37}$$

are distributed according to 'Student's' law with appropriate degrees of freedom suggests that μ and μ' are treated as sure numbers and that $\bar{x}$, $\bar{x}'$, s, and s' are random. There is no reference whatever to the variances σ^2 and σ'^2. As nothing is disclosed about what distribution they may possess, by analogy with the μ's it is natural to treat them as sure numbers also.

In order to interpret every step in calculations more easily, we shall imagine two normal populations π_1 and π_2 sampled and a sequence A of pairs of samples, of n and n' individuals respectively, drawn independently from π_1 and π_2. These pairs of samples will determine $\bar{x}$, s, $\bar{x}'$, and s', generating distributions appropriate to normal populations. Substituted into formulae (37) they will make t and t' vary to generate the two distributions of 'Student'.

With this in mind, let us examine the passage in which Fisher writes

$$\epsilon = \delta - d = s't' - st, \tag{38}$$

and comments: 'Since s' and s are known, the quantity represented on the right has a known distribution, though not one which has been fully tabulated.' We see here the same kind of sudden jump in the point of view on quantities considered as is found in the passage analysed previously. Formerly s' and s were not 'known' but random. Otherwise, the distributions of t and t' would not have been those of 'Student' but would have been normal about zero and due solely to the variability of $\bar{x}$ and $\bar{x}'$. Now s' and s are known sure numbers. Let us allow for this shift in conditions and try to visualize the character of the distribution of ϵ for fixed s' and s. For this purpose we have to consider not the whole sequence A of pairs of samples mentioned above, but only a subsequence B composed only of those pairs of samples in which the estimated variances have the same values s and s' as the ones supposed to be 'known'. The variability of ϵ in the subsequence B will be the result of the variability of $\bar{x}$ and $\bar{x}'$ only. It is known that the mean of a sample from a normal population is independent of the sample variance. Consequently

the distributions of $\bar{x}$ and $\bar{x}'$ in B will be normal. As the connexion between ϵ on one hand and $\bar{x}$ and $\bar{x}'$ on the other is linear with constant coefficients, it would follow that the distribution of ϵ in B would be normal also. Therefore, it is with some surprise that one reads Fisher's suggestion that this distribution has not been fully tabulated. Evidently, when writing the sentence quoted, Fisher had something else in mind, probably depending on the new extra-logical principle described in subsequent publications. However this may be, we have to note the conflict between the sentence quoted and the rules of ordinary logic and of the classical theory of probability.

The distribution of ϵ by itself does not play any further role in Fisher's work. Instead he and, subsequently, Sukhatme consider the ratio that we will denote by $z = \epsilon/\sqrt{(s^2+s'^2)}$. Fisher does not write any formula representing the supposed distribution of z and we have to look for the details of his ideas in Sukhatme's paper. Complimentary references to this paper in subsequent publications by Fisher suggest that it is perfectly in line with his own ideas. We quote the relevant sentence in Sukhatme's paper, only altering his notation to bring it into agreement with that of Fisher.

He (Fisher) considers the distribution of

$$z = \frac{\epsilon}{\sqrt{(s^2+s'^2)}} = t' \cos R - t \sin R, \tag{39}$$

for given n, n', and R in order to obtain the probability that z exceeds any given value.

It is obvious at once that the probability in question does not refer to either of the sequences A or B visualized above. The appropriate sequence C of pairs of samples to which this probability refers is a part of the sequence A composed of all such pairs of samples in which the variances s^2 and s'^2, while variable, keep the ratio $s/s' = \tan R =$ constant. Mathematically, the distribution sought is known as the relative distribution law of z given R and is denoted by $p(z \mid R)$. If $p(R)$ and $p(z, R)$ are the absolute probability law of R and the absolute joint probability law of z and R, respectively, then, for every R such that $p(R) > 0$,

$$p(z \mid R) = \frac{p(z, R)}{p(R)}. \tag{40}$$

The relative probability, given R, of z exceeding a fixed number z_1, that is $P(z > z_1 \mid R)$, will be obtained by integrating (40) for z from z_1 to $+\infty$. There is an alternative way of obtaining the same probability. This consists of first finding the relative joint probability law given R of t and t'. If this is denoted by $p(t, t' \mid R)$ then

$$P\{z > z_1 \mid R\} = \iint_{w(z_1)} p(t, t' \mid R)\, dt\, dt', \tag{41}$$

where the region of integration $w(z_1)$ is determined by the inequality

$$z = t' \cos R - t \sin R > z_1. \tag{42}$$

A familiar formula gives
$$p(t, t' \mid R) = \frac{p(t, t', R)}{p(R)}. \tag{43}$$

Whichever way, (40) or (43), is preferred, the resulting probability $P\{z > z_1 \mid R\}$ will have the same value and will refer to the sequence C described above.

Sukhatme has chosen to apply a quadrature procedure to calculate the integral (41) with the integrand equal to the product of two of 'Student's' distributions with $n-1$ and $n'-1$ degrees of freedom respectively. This is just the error in algebra of probability laws mentioned above. The t and t' are distributed independently and in accordance with 'Student's' laws only in the sequence A where both the means $\bar{x}$ and $\bar{x}'$ and also the variances s^2 and s'^2 are undisturbed in their random and independent variation appropriate to samples from normal populations. When calculating the probability 'for a given R', we do not consider the sequence A but only its part C so selected that the ratio s/s' is constant. This selection disturbs the original distribution of s and s' and is reflected in the resulting joint distribution of t and t'.

In our calculations above (26) we have used the letters u and v for what is here denoted by t and t'. Consequently, the joint probability law $p(t, t', R)$ is obtained from (32) by merely substituting t for u and t' for v. The absolute probability law of R is easily obtained by integrating (34) with respect to z between the limits $-\infty$ and $+\infty$. The result is

$$p(R) = c\rho^{n'-1} \frac{\sin^{n-2} R \cos^{n'-2} R}{\{n(n-1)\sin^2 R + n'(n'-1)\rho^2 \cos^2 R\}^{\frac{1}{2}(n+n'-2)}}, \tag{44}$$

with c denoting a numerical constant. Substituting (32) and (44) into (43) we obtain

$$p(t, t' \mid R) = \frac{\phi(R, \rho)}{\{n(t^2+n-1)\sin^2 R + n'(t'^2+n'-1)\rho^2 \cos^2 R\}^{\frac{1}{2}(n+n')}}, \tag{45}$$

with $\phi(R, \rho)$ denoting a function of R, ρ, n and n' only. $p(t, t' \mid R)$ is just the function to be integrated to obtain the relative probability given R of t and t' to verify any inequality such as $t' \cos R - t \sin R > z_1$. As one would expect $p(t, t' \mid R)$ appears to depend not only on R but also on the ratio of the population variances ρ^2.

It follows that, from the point of view of the ordinary theory of probability, the Fisher-Sukhatme solution is wrong. The error consists in their confusing the absolute probability law of t and t', obtainable by integrating (32) for R, with the relative probability law given R of the same variables as given by (45). Some such error seems to have been suspected by Bartlett. Repeated denials and the reference to the extra-logical principle underlying the fiducial theory lead us to believe that from the point of view of that particular theory the error is non-

existent. While accepting these explanations we may still regret that the earlier papers by Fisher and that of Sukhatme do not contain any clue as to how they are to be interpreted.

6: SUMMARY

1. The theories of fiducial argument and of confidence intervals differ in their basic conceptions. The validity of the former requires, at least in some cases, the fulfilment of various restrictions of which the theory of confidence intervals is totally free, and/or the acceptance of some new principles impossible to deduce by the rules of ordinary logic (Yates, 1939; Fisher, 1939*b*).

2. The two theories may occasionally give the same numerical results in the form of fiducial limits on one side and of confidence limits on the other. The problem of estimating the difference of means of two unknown normal populations shows, however, that this need not always be the case and that fiducial limits need not satisfy the definition of confidence limits.

3. Bartlett's criticisms of Fisher's solution of the problem just mentioned seem to be due to his considering the problem from the point of view of ordinary theory of probability and ordinary logic. In this light Fisher's solution does contain both conceptual misunderstandings (originally pointed out in the author's paper of 1934) inherent in the very concept of fiducial distribution of a parameter, and errors in algebra of probability laws. Since the first references to the new principles outside of ordinary logic, which supposedly justify the fiducial theory, were published *after* the publication of Bartlett's criticisms, the latter seem to be perfectly justified and useful.

4. Owing to a certain flaw in the ideas underlying the fiducial theory which is noticeable in passages quoted in §4, it is impossible to insist on any definite attitude towards it, except that of doubt. It may be useful, however, to express the following conjectures which seem to be very probable. If they are wrong then they will be put right and, as a result, the situation will be clarified.

The present author is inclined to think that the literature on the theory of fiducial argument was born out of ideas similar to those underlying the theory of confidence intervals. These ideas, however, seem to have been too vague to crystallize into a mathematical theory. Instead they resulted in misconceptions of 'fiducial probability' and 'fiducial distribution of a parameter' which seem to involve intrinsic inconsistencies as described in §5. In this light, the theory of fiducial inference is simply non-existent in the same sense as, for example, a theory of numbers defined by mutually contradictory definitions.

In earlier stages when the problems treated were very simple, the fallacy involved in 'fiducial probability' was not apparent. Later on, however, difficulties appeared and the new principle 'which cannot be deduced by logic' seems to have been invented to disentangle them in one particular case. But the word 'principle' implies some generality, hence the drift in comments on the same

subjects treated in 1936 and again in 1939. From the point of view of the direction of this drift it is perhaps significant that Yates speaks of 'fiducial statements' possible to make on the ground of probabilities *a posteriori* and that the paper by Jeffreys which professes the equivalence of fiducial theory with that of inverse probability appeared in the *Annals of Eugenics*, edited by R. A. Fisher.

However this may be, the only thing that the present author ventures to profess is that the theory of fiducial probability is distinct from that of confidence intervals.

REFERENCES

BARTLETT, M. S. (1936). *Proc. Camb. Phil. Soc.* **32**, 560.
—— (1939). *Ann. Math. Statist.* **10**, 129.
CLOPPER, C. J. & PEARSON, E. S. (1934). *Biometrika*, **26**, 404.
DAVID, F. N. & NEYMAN, J. (1938). *Statist. Res. Mem.* **2**, 105.
FELLER, W. (1938). *Statist. Res. Mem.* **2**, 117.
FISHER, R. A. (1925–34). *Statistical Methods for Research Workers.* London: Oliver and Boyd.
—— (1930). *Proc. Camb. Phil. Soc.* **26**, 528.
—— (1933). *Proc. Roy. Soc.* A, **139**, 343.
—— (1935). *The Design of Experiments.* London: Oliver and Boyd.
—— (1936). *Ann. Eugen., Lond.*, **6**, 391.
—— (1937). *Ann. Eugen., Lond.*, **7**, 370.
—— (1939*a*). *Ann. Eugen., Lond.*, **9**, 174.
—— (1939*b*). *Ann. Math. Statist.* **10**, 383.
FRÉCHET, M. (1937). *Recherches théoriques modernes sur la théorie des probabilités.* Paris: Gauthier-Villars.
JEFFREYS, H. (1939). *Theory of Probability.* Oxford: Clarendon Press.
—— (1940). *Ann. Eugen., Lond.*, **10**, 48.
MISES, RICHARD v. (1939). *Probability, Statistics and Truth.* London: W. Hodge and Co.
NEYMAN, J. (1934). *J. R. Statist. Soc.* **97**, 558.
—— (1935*a*). *Ann. Math. Statist.* **6**, 111.
—— (1935*b*). *Bull. Soc. Math. Fr.* **63**, 246.
—— (1937). *Philos. Trans.* A, **236**, 333.
—— (1938*a*). Lectures and Conferences on Mathematical Statistics. Graduate School, U.S. Department of Agriculture, Washington, D.C.
—— (1938*b*). *Actualités Sci. Industr.* no. 739, p. 25.
NEYMAN J. & PEARSON, E. S. (1933). *Philos. Trans.* A, **231**, 289.
PEARSON, E. S. (1939). *Biometrika*, **30**, 471.
PEARSON, KARL (1938). *The Grammar of Science.* London: Everyman's Library.
PITMAN, E. J. G. (1939). *Biometrika*, **30**, 391.
PYTKOWSKI, W. (1932). *The Dependence of Income of Small Farms upon their Area, the Outlay and the Capital Invested in Cows.* Warsaw: Series Bibljoteka Puławska, 34.
STARKEY, DAISY M. (1938). *Ann. Math. Statist.* **9**, 201.
'STUDENT' (1908). *Biometrika*, **6**, 1.
—— (1925). *Metron*, **5**, 18.
SUKHATME, P. V. (1938). *Sankhyā*, **4**, 39.
WALD, A. (1939). *Ann. Math. Statist.* **10**, 299.
WALD, A. & WOLFOWITZ, J. (1939). *Ann. Math. Statist.* **10**, 105.
WELCH, B. L. (1938). *Biometrika*, **29**, 350.
—— (1939). *Ann. Math. Statist.* **10**, 58.
YATES F. (1939). *Proc. Camb. Phil. Soc.* **35**, 579.

TABLES OF PERCENTAGE POINTS OF THE INCOMPLETE BETA-FUNCTION

COMPUTED BY CATHERINE M. THOMPSON

CONTENTS

PREFATORY NOTE

BY E. S. PEARSON

THE Incomplete Beta Function ratio has been defined as

$$I_x(p,q) = B_x(p,q)/B(p,q)$$
$$= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^x x^{p-1}(1-x)^{q-1}\,dx. \qquad (1)$$

When the fundamental *Tables of the Incomplete Beta-function* (6) were published by Karl Pearson in 1934 it was realized that they might form a basis for shorter tables suited for use in special problems. One such application is in connexion with sampling theory and the associated significance tests. In carrying out these tests it is generally considered that a table giving values of the argument corresponding to certain convenient probability levels is more useful than one in which the probability integral is listed at equal intervals of the argument. Using the transformed variable

$$z = \tfrac{1}{2}\log_e \frac{p(1-x)}{qx}, \qquad (2)$$

R. A. Fisher (3) was the first to provide tables of this character, giving values of z associated with the 0·05 and 0·01 probability levels. Since then, a table for the 0·001 level has been calculated by Colcord & Deming (2) and one for the 0·20 level by H. W. Norton (in Tables edited by Fisher & Yates (4)). In the terminology of the analysis of variance, if

S_1 is a sum of squares depending on ν_1 degrees of freedom, and

S_2 is a sum of squares depending on ν_2 degrees of freedom,

then
$$x = \frac{S_2}{S_1+S_2} \qquad (3)$$

and
$$\nu_1 = 2q, \quad \nu_2 = 2p. \qquad (4)$$

For tests of significance, z is easily computed and the fact that, when ν_1 and ν_2 are large, it tends to be normally distributed about zero with a standard deviation

$$\sigma_z = \sqrt{\left\{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}}, \tag{5}$$

lent considerable weight to its tabulation rather than that of x. Experience has, however, shown that in a number of problems it is the percentage levels of x of the Beta-distribution that are directly required; this fact and the desirability of having available a greater number of percentage levels* are reasons for the issue of the present tables giving five significant figures for x. They may be regarded as a supplement to the original 1934 *Tables* (6).

Conversion from x to z or to Snedecor's (7) F, where

$$F = \frac{\nu_2 S_1}{\nu_1 S_2} = \frac{p(1-x)}{qx}, \tag{6}$$

is straightforward. Tables of the seven percentage levels for F have, in fact, been already computed, and it is hoped to include them in a new edition of *Tables for Statisticians and Biometricians*.

Since the completion of the marginal columns for the tables of F involved some fresh computation, it seemed useful to extend the work so as to provide new tables giving thirteen percentage levels for χ^2. These tables are printed in a separate contribution on pp. 187–191 below; they have been calculated to six significant figures and cover the range of degrees of freedom $\nu = 1(1)30$ and $40(10)100$.

A word of comment is perhaps desirable as to the introduction of the notation ν, ν_1 and ν_2 for degrees of freedom in place of the customary n, n_1 and n_2. The use of the letter n, both with and without a subscript, to denote a group *frequency* has been so long established in publications associated with *Biometrika* and elsewhere that it seemed desirable in these tables to avoid confusion by adopting a fresh symbol for degrees of freedom. The letter f has sometimes been used, but the notation is not altogether satisfactory; the letter ν is that employed by Yule & Kendall ((8), p. 415), and its use here should be free from any ambiguity.

Reference has been made above to the existence of problems where the direct requirement is for the percentage levels of x rather than z or F. A case in point is that of the multiple correlation coefficient in samples from uncorrelated normally distributed material; here R^2 follows exactly the Beta distribution. In other cases the distribution may be used to give an approximate fit to probability functions whose exact equations are either unknown or difficult to handle. Thus in his Preface to the *Tables of the Incomplete Beta-function*, Karl Pearson stated that his first interest in the function was stimulated by the discovery of how accurately it could be made to graduate a hypergeometric distribution. The

* The percentage levels tabulated are: 50, 25, 10, 5, 2·5, 1 and 0·5.

fitting was carried out by equating the first four moments of the Beta and hypergeometric distributions.

Again, if a random variable can assume only values between 0 and 1, if it has a mean value of μ_1' and a second moment about zero of μ_2', then the probability law

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} x^{p-1}(1-x)^{q-1}, \tag{7}$$

where $$p = \mu_1'(\mu_1'-\mu_2')/(\mu_2'-\mu_1'^2), \quad q = (1-\mu_1')(\mu_1'-\mu_2')/(\mu_2'-\mu_1'^2), \tag{8}$$

will often give a very close approximation to the true law. Use has been made of this fact by Neyman & Pearson (5), Bishop (1) and others in determining probability levels for the likelihood ratio criterion L_1 used in testing the homogeneity of a series of variances and covariances. The accompanying tables are directly applicable in such problems.

Miss Catherine M. Thompson (now Mrs V. G. Grylls) has been responsible for by far the greater part of the numerical work involved in the production, and the tables should rightly be associated with her name. Owing to the special character of the Beta-distribution, which makes it necessary to vary the method of computation in different parts of the range of variables covered, a considerable amount of exploratory work and some careful planning was needed in the development of the lines of attack. This essential aid has been provided by Drs L. J. Comrie and H. O. Hartley of Scientific Computing Service Ltd., in whose office Miss Thompson carried out most of the work. Since the evacuation of University College at the beginning of the war this help, both in advice and in accommodation, has been more than ever essential. In the following pages Drs Comrie and Hartley have described the various methods used in computation and have also discussed the problem of interpolation.

The Editor is glad to take this opportunity of expressing his warm appreciation of this collaboration, which has made it possible to carry through to a successful conclusion a piece of work that had long been in view.

REFERENCES

(1) BISHOP, D. J. (1939). *Biometrika*, **31**, 31–55.
(2) COLCORD, C. G. & DEMING, L. S. (1935). *Sankhyā*, **2**, 423.
(3) FISHER, R. A. (1925). *Statistical Methods for Research Workers*. London: Oliver and Boyd.
(4) FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research*. London: Oliver and Boyd.
(5) NEYMAN, J. & PEARSON, E. S. (1931). *Bull. int. Acad. Cracovie*. Série A, pp. 460–81.
(6) PEARSON, K. (1934). *Tables of the Incomplete Beta-function*. London: *Biometrika*.
(7) SNEDECOR, G. W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance*. Ames, Iowa: Collegiate Press Inc.
(8) YULE, G. U. & KENDALL, M. G. (1937). *Introduction to the Theory of Statistics*. 11th ed. revised. London: C. Griffin and Co.

DESCRIPTION OF THE CALCULATION

BY L. J. COMRIE AND H. O. HARTLEY

INTRODUCTION

IN terms of Karl Pearson's notation the incomplete Beta-function $B_x(p,q)$ is defined by the integral

$$B_x(p,q) = \int_0^x x^{p-1}(1-x)^{q-1}\,dx. \tag{1}$$

For $x = 1$ we have the complete Beta-function $B_1(p,q)$, commonly denoted by $B(p,q)$ and defined by the equation

$$B(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}, \tag{2}$$

which is identical with (1) for $x = 1$.

The tables give the percentage points of the 'normalized' incomplete Beta-function

$$I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)}\int_0^x x^{p-1}(1-x)^{q-1}\,dx. \tag{3}$$

They are defined as the roots x of the equation

$$I_x(p,q) = P \tag{4}$$

for given P, as functions of the parameters p and q. Seven tables have been prepared corresponding to seven selected values of P, namely 0·005, 0·01, 0·025, 0·05, 0·10, 0·25 and 0·50. From the formula

$$I_x(p,q) = 1 - I_{1-x}(q,p) \tag{5}$$

the roots of (4) follow immediately for $P = 0{\cdot}75$, 0·90, 0·95, 0·975, 0·99 and 0·995.

In each table x is tabulated for

$$\nu_1 = 2q = 1(1)10,\ 12,\ 15,\ 20,\ 24,\ 30,\ 40,\ 60,\ 120 \text{ and } \infty$$

$$\nu_2 = 2p = 1(1)30,\ 40,\ 60,\ 120 \text{ and } \infty$$

With $2q$ as column heading and $2p$ as row headings the arrangement of the tables corresponds to that of the upper percentage points of R.A. Fisher's z and G. Snedecor's F, ν_1 and ν_2 being the degrees of freedom.

Karl Pearson in his introduction to the *Tables of the Incomplete Beta-function* (6) says: 'No single method has hitherto been discovered for evaluating numerically the incomplete Beta-function for all values of p and q.' Those who have done numerical work on this function and its various transformations will agree that the main difficulty is the limitation in scope of any single method and the variety of methods required to deal appropriately with the range of the parameters p and q and of the variable x. This difficulty is enhanced when the task is the

calculation of percentage points of x rather than the tabulation of the function $I_x(p, q)$. A large number of numerical processes, each specially designed to deal with certain ranges of the new tables, had to be employed to accomplish this task.

The choice of a suitable method is largely determined by three important factors:

(a) The accuracy required for x. This was fixed at five significant figures.

(b) Existing tables available as a starting point. These are: Karl Pearson's *Tables of the Incomplete Beta-function* (6). Fisher's tables of percentage points of z (4), and corresponding tables of its transformation F (3), and finally Karl Pearson's *Tables of the Incomplete Γ-function* (5).

(c) The number and relative position of percentage levels, P, and the values of p and q for which the percentage points x are to be calculated.

The importance of (a) and (b) is obvious, but (c) is no less relevant. It will, as a rule, be uneconomical to produce an interpolable table of $I_x(p, q)$ merely to obtain a single percentage point by inverse interpolation. If, however, a larger number of percentage levels is to be calculated the method becomes worth while. In this connexion it should be remarked that the original plan was to produce tables for $P = 0{\cdot}005$, $0{\cdot}01$, $0{\cdot}05$ and $0{\cdot}10$ only, and that it was decided at a later stage to add tables for the remaining values of P.

SUMMARY OF NUMERICAL METHODS EMPLOYED

(1) *Inverse interpolation in Karl Pearson's tables.* A large number of percentage points were obtained by inverse interpolation in the tables of $I_x(p, q)$. The particular tables required were differenced on a National machine (2) and six significant figures of x found by the method of inverse interpolation described by L. J. Comrie (2), taking into consideration the higher order differences. The method breaks down when for large p and small q (or for large q and small p) the tabular interval of $0{\cdot}01$ is too wide for the tables to be interpolable. With the notation adopted, the difficulty arises in the top right-hand corner of the tables when for large q and small p the root x is smaller than $0{\cdot}05$. Roughly speaking, roots greater than $0{\cdot}05$ could be obtained from Pearson's tables.

(2) *Interpolation in tables of percentage points.* It will be noted that in the present tables percentage points are given for certain values of p and q for which the function $I_x(p, q)$ has not been tabulated by Pearson. Such points occur in the rows $2p = 23(2)29$, 120 and in the column $2q = 120$. Whilst the calculation of the two marginal lines ($2q = 120$ and $2p = 120$) necessitated special methods, the extra entries in the interior of the table were easily obtained by p-wise interpolation, using suitable formulæ of the Lagrangian type.

(3) *Extension of the tables of $I_x(p, q)$.* The well-known recurrence formula

$$I_x(p, q) = xI_x(p-1, q) + (1-x)\,I_x(p, q-1) \tag{6}$$

is particularly convenient if $I_x(p, q)$ is required for a lattice work of integer

co-ordinates $2p$, $2q$ in a certain range and for a limited range of x. For small p and large or moderate q all percentage points of $I_x(p,q)$ are clustered near 0. It appeared worth while, therefore, to use the recurrence formula (6) to extend Pearson's table by producing

$$I_x(p,q) \quad \text{for} \quad 2p = 1(1)7, \quad 2q = 1(1)30$$

and
$$x = 0{\cdot}001(0{\cdot}001)0{\cdot}012, \ 0{\cdot}015 \text{ and } 0{\cdot}025$$

in order to obtain more of the required percentage levels x by inverse interpolation at intervals 0·005 and 0·001 respectively.* This made it possible to obtain a large number of percentage points x in a range where Pearson's tables are not interpolable.

To start the recurrence process, the functions

$$I_x(\tfrac{1}{2},q), \quad I_x(p,\tfrac{1}{2}), \quad I_x(1,q) \quad \text{and} \quad I_x(p,1)$$

are required for the above ranges of p and q. The first two functions were obtained from the expansions

$$I_x(\tfrac{1}{2},q) = \frac{1}{B(\tfrac{1}{2},q)}\left\{2x^{\frac{1}{2}} - \frac{2}{3}\frac{(q-1)}{1!}x^{\frac{3}{2}} + \frac{2}{5}\frac{(q-1)(q-2)}{2!}x^{\frac{5}{2}} - \frac{2}{7}\frac{(q-1)(q-2)(q-3)}{3!}x^{\frac{7}{2}} + \dots\right\} \tag{7}$$

$$I_x(p,\tfrac{1}{2}) = \frac{1}{B(p,\tfrac{1}{2})}\left\{\frac{x^p}{p} + \frac{1\,x^{p+1}}{2\,.\,1!(p+1)} + \frac{1\,.\,3\,x^{p+2}}{2^2\,.\,2!(p+2)} + \frac{1\,.\,3\,.\,5\,x^{p+3}}{2^3\,.\,3!(p+3)} + \dots\right\}. \tag{8}$$

In dealing with the expansion of $I_x(\tfrac{1}{2},q)$, all terms required for 10-decimal accuracy in $I_{0{\cdot}025}(\tfrac{1}{2},q)$ were first produced and then reduced for smaller values of x by multiplying by the appropriate power of $x/0{\cdot}025$. The quantities

$$I_x(1,q) = 1 - \frac{1}{q}(1-x)^q, \quad I_x(p,1) = \frac{1}{p}x^p$$

were produced with the help of logarithmic tables. The remaining functions

$$I_x(1{\cdot}5,q), \quad I_x(2{\cdot}5,q), \quad I_x(3{\cdot}5,q) \qquad \text{and} \qquad I_x(2,q), \quad I_x(3,q)$$

were then obtained by four recurrences covering the following combinations of the parameters $2p$ and $2q$.

Odd	values of $2p$ and	odd	values of $2q$
Even	,,	odd	,,
Odd	,,	even	,,
Even	,,	even	,,

Having thus dealt with the main body of the tables we now turn to the more difficult problem of calculating entries x near the margin of each table of percentage points. In what follows, methods will differ according to whether $2p$ is odd or even.

* [It is hoped that at a future date it [illegible] sh these extended values of $I_x(p,q)$ as a supplement to the *Tables of the* [illegible] Ed.]

(4) *Building up the polynomial part of $I_x(p, q)$ from a constant high-order difference* ($2p$ *even*). For integer p,

$$B_x(p,q) = \int_0^x x^{p-1}(1-x)^{q-1}\,dx, \tag{9}$$

may be expressed as a polynomial in $(1-x)$. We have

$$B(p,q) - B_x(p,q) = (1-x)^q \sum_{i=0}^{p-1} (-)^i C_i^{p-1} \frac{(1-x)^i}{q+i}, \tag{10}$$

or, introducing $y = 1-x$,

$$B_y(q,p) = y^q \sum_{i=0}^{p-1} (-)^i C_i^{p-1} \frac{y^i}{q+i}. \tag{11}$$

$B_y(q, p)$ is therefore the product of the qth power of y and a simple polynomial in y. If p is small ($2p \leqslant 12$) this polynomial can be built up easily on the National machine (2). As an example, for $2q = 60$ and $2p = 6$ we have the equation

$$14880\{B(3,30) - B_x(3,30)\} = 14880 B_y(30,3) = y^{30}(496 - 960y + 465y^2).$$

The polynomial $496 - 960y + 465y^2$ was built up on the National machine from its constant second difference for values of y beginning at $y = 1$ and descending at interval 0·001. The polynomial values were multiplied by the 30th power of the argument and the products checked by differencing. Finally the percentage points y (or x) were found by inverse interpolation. This method was used for $2q = 40$, 60, 120 and $2p = 2$, 4, 6, 8, 10, 12. For larger values of p the building up process becomes too laborious. On the other hand, with increasing p the percentage points x increase in value, so that for values of $2p$ greater than 12 and not exceeding 100 results could be obtained from Pearson's tables by inverse interpolation.

(5) *Taylor expansion at approximate percentage point* ($2p$ *even*). It remains, therefore, to consider the last column $2q = 120$ for $2p \geqslant 14$. For small x (i.e. values of y in the neighbourhood of 1) the terms of the expansion (11) have to be calculated to a very high degree of accuracy, since these terms have alternating signs and many significant figures are lost when adding to produce $B_y(q, p)$, which is very small. Since seven significant figures are required for $B_y(q, p)$, in some cases 20 decimals are required for the terms in (11), and their computation becomes laborious. A method was, therefore, evolved whereby $B_y(q, p)$ has to be calculated for one single three-decimal argument only. Although the function $B_y(q, p)$ is difficult to compute, its derivatives are easily calculated. It is, therefore, natural to use a Taylor expansion

$$B_{y+h}(q,p) - B_y(q,p) = hy^{q-1}(1-y)^{p-1}\left\{1 + \tfrac{1}{2}h\left(\frac{q-1}{y} - \frac{p-1}{1-y}\right) + \ldots\right\}. \tag{12}$$

With a known* three-decimal approximation y to the exact percentage point $y+h$ the main task consists in the calculation of $B_y(q, p)$. This was done from formula (11), using tables of powers (9), and a high capacity electric calculating

* It will be shown later how this approximation was obtained.

machine. The correction h to the approximation y was then easily obtained by iteration from equation (12). With $h_0 = 0$ the iteration is given by

$$h_{n+1} = \Delta B \Big/ \left\{1 + \tfrac{1}{2}h_n\left(\frac{q-1}{y} - \frac{p-1}{1-y}\right) + \dots\right\},$$

where
$$\Delta B = \frac{[B(p,q)\,P - B_y(q,p)]}{(1-y)^{p-1}y^{q-1}}. \tag{13}$$

Since the numerator, ΔB, of equation (13) does not vary, the corrections h_1, h_2 and h_3 are easily produced in turn, three steps being sufficient in most cases. Occasionally the term arising from the third derivative had to be included in the denominator. In this way the percentage points were calculated for

$$2p = 14(2)22 \quad 2q = 120,$$
$$2q = 14(2)22 \quad 2p = 120,$$

the values for $2q = 14, 16, 18, 22$ and $2p = 120$ being required for checking by differencing.

A word has to be added concerning the three-decimal approximation to the percentage points of x for $2q = 120$ and $2p = 120$. In some cases such values could be obtained from Fisher's table of percentage points of z using the transformation

$$x = \frac{p}{p + qe^{2z}}.$$

In cases where such values are not available they were obtained by harmonic interpolation. More precisely, the finite limits

$$\lim_{\substack{q\to\infty \\ p\,=\,\text{constant}}} 2qx \qquad \text{and} \qquad \lim_{\substack{p\to\infty \\ q\,=\,\text{constant}}} 2p(1-x)$$

were first obtained from the functions $I(u,p)$ of *Tables of the Incomplete Γ-function*. The above limits depend on p and q and are given by $2u\sqrt{p}$ and $2u\sqrt{q}$ respectively, where u is the root of $I(u,p-1) = P$ and $I(u,q-1) = 1-P$ respectively. The quantity $2qx$, being known for the arguments $1/2q = 0/120, 2/120, 3/120, 4/120, 5/120$ and $6/120$, was then obtained (to about four-decimal accuracy) for the argument $1/2q = 1/120$ by a Lagrangian interpolation formula. Similarly the quantity $2p(1-x)$ was calculated for $1/2p = 1/120$.

We are left, therefore, to consider the entries in the column $2q = 120$ with $2p$ odd, and in the row $2p = 120$ with $2q$ odd, and also certain entries in the top right-hand corner for $2q > 30$ and odd $2p < 13$.

(6) *Binomial expansion with fractional index* ($2p$ *odd*). In the top right-hand corner values of x are small and it is therefore to be expected that the expansion

$$B(p,q)\,P = \sum_{i=0}^{\infty}(-)^i\,C_i^{q-1}\frac{x^{p+i}}{p+i} \tag{14}$$

is reasonably convergent for such values of x.

The coefficients of the expansion (14) were calculated for $2q = 1(1)30$, 40, 60 and 120, and $2p = 1(1)9$. The root x of the equation (14) was then found by a suitable iteration process.

In some cases ($2p = 1, 1 \leqslant 2q \leqslant 30$) it was found convenient to invert the expansion (14). With $2p = 1$ the equation (14), if regarded as an expansion in powers of $\sqrt{x}$, may be reversed to yield $\sqrt{x}$ as an expansion in powers of $B(p, q) \times P$. Because of the particular importance of the case $2p = 1$ (the t-distribution) it is of interest to give here examples of formulæ from which any percentage point x may be obtained directly by substituting the corresponding percentage level P. If $D = \frac{1}{2}B(p,q)P = \frac{1}{2}B_x(p,q)$, the first five terms of the reversed expansion are as follows:

$$\begin{aligned}\sqrt{x} = D &+ \frac{q-1}{3}D^3 + \frac{(q-1)(7q-4)}{30}D^5 \\ &+ \frac{(q-1)(127q^2-131q+34)}{630}D^7 \\ &+ \frac{(q-1)(4369q^3-6285q^2+3042q-496)}{22680}D^9 + \dots,\end{aligned}$$

from which the expansion for any particular q may be worked out without difficulty. Thus for $q = 10$,

$$\sqrt{x} = D + 3D^3 + 19{\cdot}8D^5 + 163{\cdot}2D^7 + 1496{\cdot}2D^9 + \dots,$$

and for $q = 25$,

$$\sqrt{x} = D + 8D^3 + 136{\cdot}8D^5 + 2900{\cdot}3D^7 + 68162{\cdot}0D^9 + \dots.$$

(7) *Numerical integration.* For large p and q, when the integral $I_x(p,q)$ approaches the normal probability integral, a variety of methods has been developed (6, 7, 8). With mechanical computing aids available, numerical integration appeared to be the simplest. The integrand $x^{p-1}(1-x)^{q-1}$ represents a smooth curve and was produced at interval 0·01 with the help of logarithmic tables and checked by differencing on the National machine. Numerical integration was performed by Gauss' formula and the integral $B_x(p,q)$ checked by differencing. Finally, x was obtained by inverse interpolation and checked by the application of Taylor's expansion at the tabular value nearest to x. This method was used for $2q = 120$ and $2p = 24$, 26, 28, 30, 40, 60 and 120 and also for $2p = 120$ and $2q = 24$, 30, 40 and 60.

For $2q = 120$ and $2p = 7(2)29$ all percentage points were obtained by p-wise interpolation between the entries $2p = 2(2)30$, 40 and 60, using appropriate formulæ of the Lagrangian type.

(8) *Approximation by the incomplete Γ-function.* It remains to consider the entries for

$$2p = 120 \quad \text{and} \quad 2q = 1(2)9.$$

There appears to be a lack of suitable methods for obtaining accurate results in this range of q for isolated large values of p. Three-decimal approximations x_0 to

the percentage levels may be obtained by harmonic interpolation as described in § (5).* To obtain the correct percentage points ($x = x_0 + h$), the main task consists in calculating $I_{x_0}(60, q)$ to six places of decimals. This was done with the help of a recently developed approximate formula giving $I_x(p, q)$ in terms of the incomplete Γ-function.† This formula is akin to a Taylor expansion of $I_x(p, q)$ in powers of $1/2p$ at $1/2p = 0$ $(2p = \infty)$ and may be written as follows:

$$1 - I_x(p, q) \cong I(u, q-1) + \frac{e^{-\lambda}\lambda^q}{\Gamma(q)}\left\{\frac{T_1}{2p} + \frac{T_2}{(2p)^2} + \frac{T_3}{(2p)^3} + \dots\right\}, \tag{15}$$

where
$$u = \frac{p(1-x)}{x\sqrt{q}} \quad \text{and} \quad \lambda = \frac{p(1-x)}{x}$$

and the terms T_i are dependent on λ and q only, the first two being

$$T_1 = q - 1 - \lambda$$
$$T_2 = \tfrac{1}{2}q^3 - \tfrac{5}{3}q^2 + \tfrac{3}{2}q - \tfrac{1}{3} + (-\tfrac{3}{2}q^2 + \tfrac{11}{6}q - \tfrac{1}{3})\lambda + (\tfrac{3}{2}q - \tfrac{1}{6})\lambda^2 - \tfrac{1}{2}\lambda^3.$$

This formula is very accurate for large p and small or moderate q. The terms T_1 and T_2 were calculated in each case for

$$\lambda = \lambda_0 = \frac{p(1-x_0)}{x_0},$$

where x_0 denotes a suitable three-decimal approximation to the true percentage point. To obtain T_3 we make use of the fact that equation (15) should yield $I_x(50, q)$ to seven-decimal accuracy. Since $I_x(50, q)$ is obtained to that accuracy from Pearson's table and since T_1, T_2 and T_3 depend on λ and q only, we may use equation (15) to determine T_3 by substituting $p = 50$, $\lambda = \lambda_0$, $x = x_1 = \dfrac{p}{\lambda_0 + p}$, $u = \dfrac{p(1-x_1)}{x_1\sqrt{q}}$. With T_1, T_2 and T_3 computed, $I_{x_0}(60, q)$ is easily obtained from equation (15). Finally the exact percentage point $x_0 + h$ is calculated by the iteration process (13).

Checks

The main body of each table of percentage points (i.e. the interior of each table) was checked by differencing p-wise and q-wise at interval $\frac{1}{2}$. For large q and moderate p, x may be considered as a function of $1/2q$ and differenced at interval $1/120$, i.e. for $1/2q = 0/120\ (1/120)\ 6/120$. Similarly x may be differenced for large p and moderate values of q for $1/2p = 0/120\ (1/120)\ 6/120$. Four significant figures may be checked in this way, thus eliminating any possibility of serious errors. For small p and large q, the quantity x is almost linear in $1/2q$ so that a good check was given by examination of the product $2qx$, which is almost constant. Never-

* In this case it was sufficient to use a Lagrangian formula for interpolation between values of x (with $x = 1$ for $2p = \infty$), instead of performing the more complicated interpolation between values of $2p(1-x)$.

† The derivation of this formula is given in a paper by H. O. Hartley which will, it is hoped, be published in the next issue of *Biometrika*.

theless, the only available check to guarantee five-decimal accuracy at the margins was recomputation. As far as possible repetition of the method employed in the first instance was avoided. Thus inverse interpolation was replaced by direct interpolation or by a Taylor expansion at a tabular value; iteration processes were varied in the formulæ employed.

METHODS OF INTERPOLATION

BY H. O. HARTLEY

INTRODUCTION

IN so far as the table is required in connexion with standard tests of significance the user will be concerned with obtaining x for any percentage level P and for integer values of $\nu_2 = 2p$ and $\nu_1 = 2q$.

The values of P and the row and column headings $(2p, 2q)$ have been selected in such a way that the user will generally find the required value of x tabulated. Moreover, for most of the applications it suffices to estimate roughly the magnitude of interpolates from an inspection of the table. In some cases, however, interpolation to about five-decimal accuracy is necessary. The problem of interpolating between corresponding entries in different tables of percentage points (interpolation P-wise) will, it is hoped, be dealt with elsewhere and we are here only concerned with interpolation in each individual table of percentage points $x(2p, 2q)$ to find x for any combination of integer arguments $2p$, $2q$.*

In the present tables interpolation to integer arguments $\nu_2 = 2p$, $\nu_1 = 2q$ will occur in three different forms:

(1) Single-entry interpolation q-wise in the range $1 \leqslant 2p \leqslant 30$ and $10 \leqslant 2q < \infty$.

(2) Single-entry interpolation p-wise in the range $1 \leqslant 2q \leqslant 10$ and $30 \leqslant 2p < \infty$.

(3) Double-entry interpolation for $30 \leqslant 2p$, $10 \leqslant 2q$.

If both $2p$ and $2q$ are large, interpolation in the tables is impractical, and it was therefore necessary to add a fourth section, namely:

(4) Approximate calculation of percentage points if both p and q are large.

It will be noted that, following the lay-out adopted in other tables of percentage points (3, 4), the column headings $\nu_1 = 2q = 20, 24, 30, 40, 60, 120$ and ∞ are in harmonic progression. If, therefore, $1/2q$ is used as a variable, these columns form a tabulation of x at equidistant intervals of the variable $1/2q$. The same harmonic progression is given for the row headings $\nu_2 = 2p$, although here the tabulation at unit interval goes up to $2p = 30$, because of

* Fractional values of $2p$ and $2q$ occur in a number of applications when the percentage points of certain Pearson-type curves are required. In such cases it will be found most convenient to apply single-entry interpolation formulæ, first in one direction (p or q) and then in the other. Methods akin to those given here cover the range $2q > 10$, $2p > 30$. For the range $0 \leqslant 2p \leqslant 30$, $0 \leqslant 2q \leqslant 10$, successive single-entry interpolation (p-wise or q-wise) at unit interval should afford no difficulty provided the arguments of the interpolate do not lie within the strips $0 \leqslant 2p \leqslant 3$, $0 \leqslant 2q \leqslant 2$. Within these strips interpolation cannot be carried out without the aid of auxiliary tables.

the importance of these values for certain tests of significance. The use of the variables $1/2q$ and $1/2p$ greatly facilitates interpolation but even with this device (known as harmonic interpolation) high-order interpolation formulæ have to be used in many cases, if the accuracy of the tabular values is required.

To facilitate single-entry interpolation, therefore, an auxiliary table of Lagrangian coefficients has been prepared. Although this auxiliary table has been specifically designed to meet the requirements of the present tables of percentage points of x, it is given and described in a separate paper (p. 183) since it is felt that it will have a wider application to any table of percentage points with a similar lay-out.

1. Single-entry interpolation q-wise

No interpolation is required for the range $1 \leqslant 2q \leqslant 10 (1 \leqslant \nu_1 \leqslant 10)$. For $\nu_1 = 2q > 10$ interpolates are obtained with the help of the auxiliary table on pp. 183–5 of this issue, and its use is best explained in terms of an example.

Example 1. Find the 5 % point corresponding to $2p = 26$, $2q = 74$.

In the auxiliary table (p. 185 below) enter the row headed 74, that is, the row whose heading is equal to the value of $2q$ for which the interpolate is required. The entries in this row are the (Lagrangian) multipliers in a sum of products which yields the interpolate x. The corresponding multiplicands are taken from the table of 5 % points x $(2p, 2q)$. We enter the row headed $2p = 26$ and select entries $x(26, 2q)$ for $2q = 20$, 24, 30, 40, 60, 120 and ∞. These correspond to the column headings in the auxiliary table. The sign of each product is also given at the top of the columns. We have, therefore,

$$\begin{aligned} x(26, 74) &= +0{\cdot}395\,16 \times 0{\cdot}005\,867 - 0{\cdot}357\,56 \times 0{\cdot}045\,623 + 0{\cdot}313\,14 \times 0{\cdot}162\,013 \\ &\quad - 0{\cdot}259\,66 \times 0{\cdot}372\,737 + 0{\cdot}193\,79 \times 1{\cdot}018\,370 + 0{\cdot}110\,24 \times 0{\cdot}247\,951 \\ &= 0{\cdot}164\,64. \end{aligned}$$

This may be compared with the exact value 0·164 637 obtained by inverse interpolation from Pearson's tables.

The accuracy of the interpolates depends on $2p$ and $2q$ and (to a lesser extent) on the percentage level P. In favourable cases, if $2p$ is small and $2q$ moderate, the interpolate is accurate to 5 places of decimals. In the least favourable cases, for $2p$ near 30 or $2q$ large, the fifth decimal of the interpolate may be in error. One more example is given to demonstrate the use of the auxiliary table.

Example 2. Find the 50 % point for $2p = 11$, $2q = 17$.

Entering the row 17 in the auxiliary table and the row $2p = 11$ in the table of 50 % points we have

$$\begin{aligned} x(11, 17) &= +0{\cdot}525\,38 \times 0{\cdot}003\,097 - 0{\cdot}476\,96 \times 0{\cdot}037\,459 + 0{\cdot}419\,02 \times 0{\cdot}409\,711 \\ &\quad + 0{\cdot}348\,45 \times 1{\cdot}213\,958 - 0{\cdot}307\,07 \times 0{\cdot}856\,212 + 0{\cdot}260\,64 \times 0{\cdot}315\,162 \\ &\quad - 0{\cdot}208\,18 \times 0{\cdot}048\,257 \\ &= 0{\cdot}387\,62. \end{aligned}$$

It will be noted that for $2q = 16$, 17, 18 and 19 two alternative rows are given in the auxiliary table, one (which has been used in the above example) is under the heading 'Harmonic' Lagrangian coefficients. The other row contains 'Ordinary' Lagrangian coefficients. It is in this range of $2q$ that there is little to choose between the merits of ordinary and harmonic interpolation, and the use of both methods provides a good check. Reworking the above example and using ordinary Lagrangian coefficients we have

$$\begin{aligned} x(11, 17) &= -0{\cdot}553\,46 \times 0{\cdot}306\,397 + 0{\cdot}525\,38 \times 0{\cdot}780\,000 - 0{\cdot}476\,96 \times 0{\cdot}983\,025 \\ &\quad + 0{\cdot}419\,02 \times 1{\cdot}258\,272 + 0{\cdot}348\,45 \times 0{\cdot}289\,546 - 0{\cdot}307\,07 \times 0{\cdot}040\,124 \\ &\quad + 0{\cdot}260\,64 \times 0{\cdot}001\,728 \\ &= 0{\cdot}387\,62. \end{aligned}$$

There is satisfactory agreement between the two interpolates and the exact value (obtained from Pearson's table) which is 0·387 619.

2. SINGLE-ENTRY INTERPOLATION p-WISE

No interpolation is required for $1 \leqslant 2p \leqslant 30$ $(1 \leqslant \nu_2 \leqslant 30)$. For $\nu_2 = 2p > 30$ we again use the auxiliary table on pp. 184, 185 of this issue. This time, however, the argument $2p$ of the interpolate determines the row to be entered in the auxiliary table, whilst column headings of this table are made to correspond to selected *rows* in the table of percentage points. The method is best explained by the following examples.

Example 3. Find the 0·5 % point corresponding to $2q = 4$ and $2p = 96$. In the auxiliary table enter the row headed 96. The entries in this row are the Lagrangian multipliers. The corresponding multiplicands are taken from the table of 0·5 % points. We enter the column headed $2q = 4$ and select entries $x(2p, 4)$ for $2p = 20$, 24, 30, 40, 60, 120 and ∞ which correspond to the column heading in the auxiliary table.

The sign of each product is also given at the top of the columns. We therefore have

$$\begin{aligned} x(96, 4) &= +0{\cdot}491\,44 \times 0{\cdot}005\,875 - 0{\cdot}550\,98 \times 0{\cdot}044\,647 + 0{\cdot}618\,64 \times 0{\cdot}152\,207 \\ &\quad - 0{\cdot}695\,71 \times 0{\cdot}318\,909 + 0{\cdot}783\,70 \times 0{\cdot}558\,090 + 0{\cdot}884\,42 \times 0{\cdot}669\,708 \\ &\quad - 1{\cdot}000\,00 \times 0{\cdot}022\,324 \\ &= 0{\cdot}857\,94, \end{aligned}$$

which differs by 2 units in the fifth decimal from the exact value (0·857 92) obtained by inverse interpolation from Pearson's tables.

Example 4. Find the 5 % point corresponding to $2p = 80$ and $2q = 30$. We have

$$\begin{aligned} x(80, 30) &= +0{\cdot}246\,39 \times 0{\cdot}006\,836 - 0{\cdot}292\,08 \times 0{\cdot}052\,734 + 0{\cdot}352\,00 \times 0{\cdot}184\,570 \\ &\quad - 0{\cdot}433\,21 \times 0{\cdot}410\,156 + 0{\cdot}548\,07 \times 0{\cdot}922\,851 + 0{\cdot}720\,16 \times 0{\cdot}369\,141 \\ &\quad - 1{\cdot}000\,00 \times 0{\cdot}020\,508 \\ &= 0{\cdot}624\,69. \end{aligned}$$

The exact value is 0·624 75.

Again, the accuracy of the interpolates depends on $2q$, $2p$ and to a lesser extent on P. Five-decimal accuracy is obtained for small $2q$ and moderate $2p$, whilst only 4 decimals are reliable if $2q$ is near 30 or $2p$ is large.

3. Harmonic double-entry interpolation

In this section we deal with interpolation in the range

$$30 \leqslant 2p < \infty, \quad 10 \leqslant 2q < \infty,$$

provided the arguments of the interpolate are not 'too large'. The exact meaning

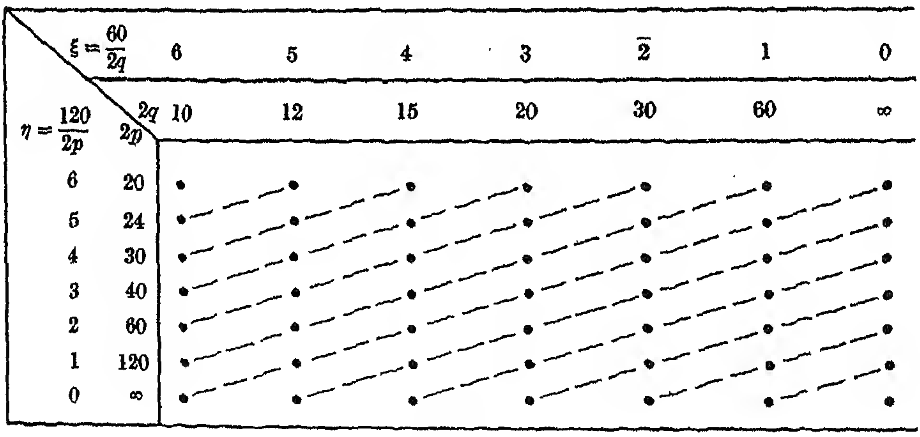


Fig. 1.

of this restriction is that if the methods described below are used to find interpolates in the range

$$60 \leqslant 2p < \infty, \quad 40 \leqslant 2q < \infty,$$

the results obtained are unsatisfactory. In this range the user should, therefore, proceed on lines described in section 4.

The method is essentially double-entry interpolation between points of the lattice work shown in Fig. 1.

The 5, $2\frac{1}{2}$, 1 and $\frac{1}{2}$ % values (i.e. the quantities x for $P = 0{\cdot}05$, $0{\cdot}025$, $0{\cdot}01$ and $0{\cdot}005$) are practically linear to three-figure accuracy in the diagonal direction indicated by the broken lines in Fig. 1.

To explain the method it will be convenient to regard the percentage points x as functions of

$$\eta = \frac{120}{2p} \quad \text{and} \quad \xi = \frac{60}{2q},$$

and to introduce the notation

$$x[\eta, \xi] = x(2p, 2q).$$

The relation between the argument η, ξ and $2p$, $2q$ is demonstrated in Fig. 1. To obtain the interpolate x for any p, q in the above range calculate

$$\xi = \frac{60}{2q} \quad \text{and} \quad \eta = \frac{120}{2p}$$

and find

$$\begin{aligned} \Xi &= \text{integral part of } \xi, \\ H &= \text{integral part of } \eta, \\ \mu &= (\xi - \Xi) + (\eta - H) - 1. \end{aligned}$$

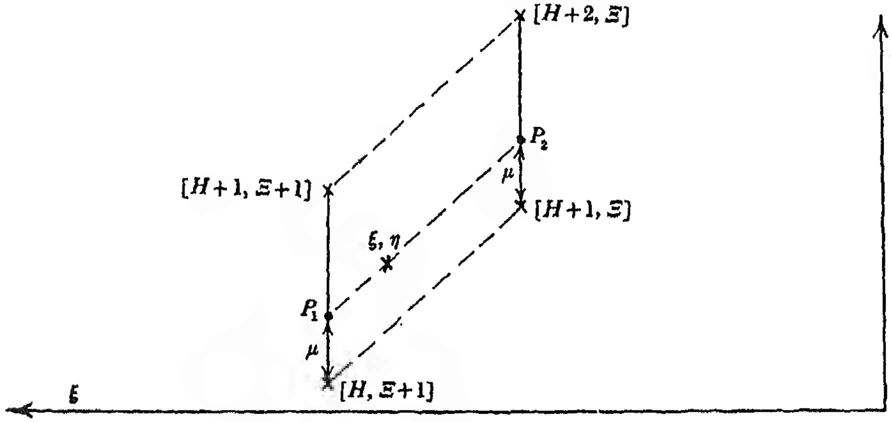


Fig. 2.

If μ is positive consider the parallelogram with vertices at $[H, \Xi+1]$, $[H+1, \Xi]$, $[H+1, \Xi+1]$ and $[H+2, \Xi]$ (see Fig. 2). Now calculate two interpolates x_1 and x_2 at points P_1 and P_2 from the (approximate) formulæ

$$\left.\begin{aligned} x_1 &= \mu x[H+1, \Xi+1] + (1-\mu)\, x[H, \Xi+1] \\ x_2 &= \mu x[H+2, \Xi] + (1-\mu)\, x[H+1, \Xi] \end{aligned}\right\} \tag{16}$$

and finally find the interpolate $x[\eta, \xi]$

$$x[\eta, \xi] = (\xi - \Xi)\, x_1 + (\Xi + 1 - \xi)\, x_2. \tag{17}$$

If μ is negative the points P_1 and P_2 will be at distance μ below the points $[H, \Xi+1]$ and $[H+1, \Xi]$ respectively, and we have the formulæ

$$\begin{aligned} x_1 &= -\mu x[H-1, \Xi+1] + (1+\mu)\, x[H, \Xi+1] \\ x_2 &= -\mu x[H, \Xi] + (1+\mu)\, x[H+1, \Xi] \end{aligned}$$

in place of equations (16).

Example 5. Find the 1 % point x for $2p = 42$, $2q \doteq 16$.

We have

$$\xi = 3{\cdot}75, \quad \eta = 2{\cdot}8571, \quad \Xi = 3, \quad H = 2, \quad \mu = 0{\cdot}6071.$$

To apply formulæ (16) the tabular values are taken from the table of 1 % points, where we find to 4-decimal accuracy

$$x[3,4] = x(40,15) = 0{\cdot}5140, \quad x[4,3] = x(30,20) = 0{\cdot}3705,$$
$$x[2,4] = x(60,15) = 0{\cdot}6297, \quad x[3,3] = x(40,20) = 0{\cdot}4578.$$

Applying the equations (16) we obtain

$$x_1 = 0{\cdot}6071 \times 0{\cdot}5140 + 0{\cdot}3929 \times 0{\cdot}6297 = 0{\cdot}5595$$
$$x_2 = 0{\cdot}6071 \times 0{\cdot}3705 + 0{\cdot}3929 \times 0{\cdot}4578 = 0{\cdot}4048.$$

Finally we calculate

$$x[2{\cdot}857, 3{\cdot}75] = x(42,16) = 0{\cdot}75 \times 0{\cdot}5595 + 0{\cdot}25 \times 0{\cdot}4048 = 0{\cdot}5208.$$

The exact value obtained from Pearson's table is $0{\cdot}5163$. If higher accuracy is required we have to improve the approximate relations (16) by adding the second-order difference effect. If this is done we obtain

$$x_1 = 0{\cdot}5555, \quad x_2 = 0{\cdot}4016 \quad \text{and} \quad x = 0{\cdot}5170,$$

which agrees satisfactorily with the exact value. If this method is applied to the tables of 10, 25 and 50 % points and if a similar precision is required, the second-order difference effect should also be considered when interpolating along the diagonals. In such cases the right-hand side of equation (17) should have four terms.

4. Calculation of percentage points if both p and q are large

If x is required for values of $2p$ and $2q$ in the range $2p > 60$, $2q > 40$, interpolation between the tabular values is not possible because of the singularity of x at $2p = \infty$, $2q = \infty$. In this range therefore x has to be calculated *ab initio*. Certain approximate formulæ for the incomplete beta-function are valid in this range(8, 10). These formulæ, whilst useful for a calculation of $P = I_x(p, q)$ as functions of x, $2p$ and $2q$, cannot be easily inverted to yield x for the given percentage levels P.

Auxiliary table

y = normal deviate at level P, $\lambda = \frac{1}{6}(y^2+3)$

P	0·50	0·25	0·10	0·05	0·025	0·01	0·005
y	0·0000	0·6745	1·2816	1·6449	1·9600	2·3263	2·5758
λ	0·5000	0·5758	0·7737	0·9509	1·1402	1·4020	1·6058
$\lambda-\frac{1}{6}$	0·3333	0·4092	0·6071	0·7843	0·9736	1·2353	1·4392

A more convenient approximation of sufficient accuracy has recently been given by Cochran(1), who extended a method suggested by Fisher(4). It is essentially an approximation (by the normal distribution) to Fisher's z-trans-

formation of x and it involves values of the normal deviate y at the appropriate percentage levels P. These normal deviates y together with a function of y denoted by λ are tabulated above for the levels P with which we are concerned here.

To find an approximation to x for any pair of arguments $2p$, $2q$ in the above range, calculate in turn the quantities

$$A = \frac{8pq}{2p+2q}$$

$$z = \frac{y}{\sqrt{(A-\lambda)}} + \frac{(\lambda - \frac{1}{6})(A-2p)}{pA}$$

$$x = \frac{2p}{2p+2q\,e^{2z}}.$$

As examples we consider two tabular values of x in order to obtain some idea of the accuracy of the approximation.

Example 6. $P = 0{\cdot}01,\quad 2p = 120,\quad 2q = 40,$

$$A = \frac{8\times 60\times 20}{160} = 60,\quad z = \frac{2{\cdot}3263}{\sqrt{58{\cdot}598}} + \frac{1{\cdot}235(-60)}{60\times 60} = 0{\cdot}2833,\quad x = 0{\cdot}6299.$$

This agrees with the exact value to four decimals.

Example 7. $P = 0{\cdot}50,\quad 2p = 30,\quad 2q = 120,$

$$A = \frac{8\times 15\times 60}{150} = 48,\quad z = \frac{0{\cdot}333\times 18}{15\times 48} = 0{\cdot}00833,\quad x = 0{\cdot}1974.$$

This differs from the exact value by about a unit in the fourth decimal.

REFERENCES

(1) COCHRAN, W. G. (1940). *Ann. Math. Statist.* **11**, 93.
(2) COMRIE, L. J. (1936). *J.R. Statist. Soc.* **3**, 87.
(3) FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological Agricultural and Medical Research.* Oliver and Boyd.
(4) FISHER, R. A. (1936). *Statistical Methods for Research Workers*, p. 237. Edinburgh: Oliver and Boyd.
(5) PEARSON, K. (1922). *Tables of the Incomplete Γ-function.* London: *Biometrika.*
(6) —— (1934). *Tables of the Incomplete Beta-function.* London: *Biometrika.*
(7) —— (1931). *Tables for Statisticians and Biometricians*, Part II. London: *Biometrika.*
(8) SOPER, H. E. (1921). *Tracts for Computers*, No. 7. Cambridge University Press.
(9) U.S.A. Work Program W.P.A. (1939). *Table of the First Ten Powers of the Integers from* 1 *to* 1000.
(10) WISHART, J. (1927). *Biometrika*, **19**, 1.

Beta Distribution: 50 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	0·50000	0·25000	0·16319	0·12061	0·095526	0·079033	0·067378	0·058711	0·052015
2	·75000	·50000	·37004	·29289	·24214	·20630	·17966	·15910	·14276
3	·83681	·62996	·50000	·41363	·35245	·30695	·27181	·24386	·22112
4	·87939	·70711	·58637	·50000	·43556	·38573	·34609	·31381	·28703
5	0·90447	0·75786	0·64755	0·56444	0·50000	0·44867	0·40684	0·37213	0·34286
6	·92097	·79370	·69305	·61427	·55133	·50000	·45737	·42141	·39068
7	·93262	·82034	·72819	·65391	·59316	·54263	·50000	·46355	·43205
8	·94129	·84090	·75614	·68619	·62787	·57859	·53645	·50000	·46818
9	·94799	·85724	·77888	·71297	·65714	·60932	·56795	·53182	·50000
10	0·95331	0·87055	0·79775	0·73555	0·68214	0·63538	0·59546	0·55984	0·52824
11	·95765	·88159	·81366	·75484	·70376	·65907	·61968	·58471	·55346
12	·96125	·89090	·82725	·77151	·72262	·67948	·64116	·60692	·57613
13	·96429	·89885	·83899	·78606	·73923	·69759	·66035	·62687	·59661
14	·96689	·90572	·84924	·79887	·75396	·71376	·67760	·64490	·61520
15	0·96913	0·91172	0·85827	0·81023	0·76712	0·72830	0·69318	0·66127	0·63216
16	·97109	·91700	·86627	·82038	·77894	·74143	·70732	·67620	·64768
17	·97282	·92169	·87342	·82950	·78963	·75334	·72022	·68986	·66195
18	·97435	·92587	·87985	·83774	·79932	·76421	·73203	·70242	·67511
19	·97572	·92964	·88565	·84522	·80817	·77417	·74288	·71401	·68728
20	0·97695	0·93303	0·89092	0·85204	0·81626	0·78331	0·75289	0·72472	0·69858
21	·97806	·93612	·89573	·85828	·82370	·79175	·76215	·73467	·70909
22	·97907	·93893	·90013	·86402	·83057	·79955	·77074	·74392	·71889
23	·97999	·94151	·90417	·86931	·83692	·80679	·77873	·75254	·72805
24	·98083	·94387	·90790	·87421	·84281	·81353	·78618	·76061	·73663
25	0·98161	0·94606	0·91135	0·87875	0·84828	0·81981	0·79315	0·76816	0·74469
26	·98232	·94808	·91455	·88298	·85340	·82568	·79968	·77526	·75227
27	·98298	·94995	·91753	·88692	·85817	·83118	·80581	·78193	·75941
28	·98360	·95170	·92031	·89060	·86265	·83635	·81157	·78821	·76615
29	·98417	·95332	·92290	·89406	·86685	·84120	·81701	·79415	·77253
30	0·98470	0·95484	0·92534	0·89730	0·87080	0·84578	0·82214	0·79976	0·77856
40	·98855	·96594	·94324	·92136	·90038	·88030	·86107	·84266	·82501
60	·99238	·97716	·96164	·94645	·93166	·91731	·90338	·88985	·87672
120	·99620	·98851	·98056	·97264	·96482	·95710	·94951	·94202	·93465
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}50$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 50 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	10	12	15	20	24	30	40	60	120
1	0·046687	0·038746	0·030867	0·023052	0·019168	0·015301	0·011450	0·0076165	0·0037997
2	·12945	·10910	·088278	·066967	·056126	·045158	·034064	·022840	·011486
3	·20225	·17275	·14173	·10908	·092099	·074664	·056756	·038355	·019443
4	·26445	·22849	·18977	·14796	·12579	·10270	·078644	·053552	·027361
5	0·31786	0·27738	0·23288	0·18374	0·15719	0·12920	0·099622	0·068335	0·035184
6	·36412	·32052	·27170	·21669	·18647	·15422	·11970	·082690	·042896
7	·40454	·35884	·30682	·24711	·21382	·17786	·13893	·096624	·050494
8	·44016	·39308	·33873	·27528	·23939	·20024	·15734	·11015	·057977
9	·47176	·42387	·36784	·30142	·26337	·22144	·17499	·12328	·065345
10	0·50000	0·45169	0·39451	0·32575	0·28589	0·24154	0·19192	0·13603	0·072602
11	·52538	·47696	·41902	·34845	·30707	·26064	·20818	·14842	·079747
12	·54831	·50000	·44162	·36967	·32704	·27880	·22379	·16046	·086785
13	·56912	·52110	·46254	·38956	·34589	·29610	·23880	·17217	·093716
14	·58811	·54049	·48194	·40823	·36371	·31258	·25325	·18355	·10054
15	0·60549	0·55838	0·50000	0·42579	0·38059	0·32832	0·26715	0·19463	0·10727
16	·62147	·57492	·51684	·44234	·39660	·34335	·28055	·20541	·11390
17	·63621	·59027	·53258	·45797	·41181	·35772	·29347	·21591	·12043
18	·64984	·60456	·54733	·47274	·42626	·37147	·30593	·22613	·12686
19	·66248	·61788	·56118	·48673	·44002	·38465	·31796	·23609	·13320
20	0·67425	0·63033	0·57421	0·50000	0·45314	0·39729	0·32958	0·24580	0·13945
21	·68522	·64200	·58649	·51260	·46566	·40942	·34082	·25527	·14561
22	·69548	·65295	·59807	·52458	·47762	·42108	·35168	·26450	·15168
23	·70509	·66325	·60903	·53599	·48905	·43228	·36219	·27350	·15767
24	·71411	·67296	·61941	·54686	·50000	·44305	·37237	·28229	·16357
25	0·72260	0·68213	0·62924	0·55723	0·51049	0·45343	0·38222	0·29086	0·16939
26	·73060	·69079	·63859	·56714	·52054	·46342	·39177	·29924	·17513
27	·73815	·69900	·64747	·57662	·53019	·47306	·40103	·30741	·18079
28	·74529	·70678	·65593	·58569	·53946	·48236	·41001	·31540	·18638
29	·75205	·71417	·66399	·59437	·54837	·49133	·41873	·32321	·19189
30	0·75846	0·72120	0·67168	0·60271	0·55695	0·50000	0·42720	0·33084	0·19732
40	·80808	·77621	·73285	·67042	·62763	·57280	·50000	·39866	·24791
60	·86397	·83954	·80537	·75420	·71771	·66916	·60134	·50000	·33209
120	·92740	·91321	·89273	·86055	·83643	·80268	·75209	·66791	·50000
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

Beta Distribution: 25 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9
1	0·14645	0·062500	0·039063	0·028309	0·022173	0·018215	0·015453	0·013416	0·011853
2	·43750	·25000	·17452	·13397	·10870	·091440	·078908	·069395	·061929
3	·59715	·39685	·29801	·23885	·19937	·17113	·14991	·13339	·12015
4	·68878	·50000	·39448	·32635	·27852	·24302	·21560	·19376	·17596
5	0·74711	0·57435	0·46936	0·39775	0·34546	0·30550	0·27390	0·24828	0·22707
6	·78726	·62996	·52848	·45632	·40198	·35944	·32516	·29692	·27323
7	·81650	·67295	·57609	·50494	·45001	·40614	·37021	·34022	·31478
8	·83872	·70711	·61516	·54582	·49117	·44680	·40996	·37885	·35219
9	·85616	·73487	·64773	·58060	·52678	·48245	·44521	·41343	·38597
10	0·87021	0·75786	0·67529	0·61052	0·55783	0·51390	0·47662	0·44451	0·41655
11	·88177	·77720	·69888	·63651	·58513	·54184	·50475	·47257	·44435
12	·89144	·79370	·71931	·65929	·60930	·56679	·53009	·49801	·46970
13	·89966	·80793	·73716	·67941	·63085	·58921	·55300	·52116	·49289
14	·90672	·82034	·75288	·69730	·65017	·60946	·57382	·54230	·51419
15	0·91285	0·83124	0·76684	0·71332	0·66758	0·62782	0·59282	0·56169	0·53380
16	·91823	·84090	·77932	·72773	·68336	·64456	·61021	·57953	·55192
17	·92298	·84951	·79053	·74077	·69772	·65986	·62619	·59599	·56870
18	·92721	·85724	·80066	·75263	·71084	·67391	·64093	·61122	·58428
19	·93100	·86422	·80986	·76345	·72287	·68686	·65456	·62536	·59879
20	0·93442	0·87055	0·81825	0·77337	0·73395	0·69882	0·66720	0·63852	0·61234
21	·93751	·87632	·82593	·78250	·74418	·70991	·67895	·65079	·62500
22	·94033	·88159	·83299	·79092	·75366	·72021	·68991	·66226	·63688
23	·94290	·88644	·83950	·79871	·76246	·72981	·70015	·67300	·64803
24	·94526	·89090	·84553	·80595	·77066	·73878	·70973	·68309	·65852
25	0·94744	0·89503	0·85112	0·81268	0·77831	0·74717	0·71873	0·69258	0·66840
26	·94944	·89885	·85632	·81896	·78547	·75505	·72719	·70151	·67774
27	·95130	·90241	·86116	·82484	·79218	·76244	·73515	·70995	·68657
28	·95303	·90572	·86570	·83035	·79848	·76941	·74267	·71793	·69492
29	·95464	·90882	·86994	·83552	·80442	·77598	·74977	·72548	·70285
30	0·95614	0·91172	0·87393	0·84039	0·81002	0·78219	0·75649	0·73263	0·71038
40	·96706	·93303	·90351	·87685	·85230	·82947	·80809	·78797	·76896
60	·97801	·95484	·93434	·91548	·89782	·88113	·86525	·85009	·83557
120	·98899	·97716	·96648	·95647	·94692	·93774	·92887	·92025	·91187
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}25$ where $p = \tfrac{1}{2}\nu_2$, $q = \tfrac{1}{2}\nu_1$.

BETA DISTRIBUTION: 25 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120
1	0·010616	0·0087814	0·0069734	0·0051914	0·0043101	0·0034353	0·0025669	0·0017049	$0·0^3 84926$
2	·055913	·046816	·037631	·028358	·023689	·018996	·014281	·0095436	·0047832
3	·10930	·092592	·075324	·057467	·048307	·038986	·029500	·019844	·010012
4	·16116	·13797	·11350	·087610	·074095	·060174	·045827	·031031	·015764
5	0·20922	0·18082	0·15026	0·11726	0·099749	0·081498	0·062458	0·042571	0·021774
6	·25307	·22058	·18500	·14585	·12475	·10251	·079043	·054222	·027923
7	·29291	·25724	·21758	·17316	·14888	·12301	·095405	·065857	·034143
8	·32908	·29099	·24802	·19913	·17203	·14289	·11145	·077403	·040396
9	·36198	·32205	·27644	·22376	·19420	·16211	·12712	·088814	·046656
10	0·39196	0·35068	0·30297	0·24710	0·21538	0·18064	0·14240	0·10006	0·052904
11	·41938	·37712	·32776	·26921	·23562	·19850	·15726	·11113	·059127
12	·44451	·40158	·35094	·29017	·25493	·21570	·17171	·12200	·065317
13	·46762	·42426	·37265	·31004	·27338	·23226	·18575	·13267	·071466
14	·48893	·44534	·39302	·32889	·29100	·24819	·19938	·14314	·077570
15	0·50863	0·46497	0·41215	0·34679	0·30785	0·26353	0·21261	0·15341	0·083624
16	·52691	·48330	·43016	·36380	·32395	·27831	·22546	·16347	·089625
17	·54389	·50043	·44712	·37998	·33936	·29254	·23793	·17332	·095571
18	·55972	·51649	·46312	·39539	·35411	·30625	·25004	·18298	·10146
19	·57449	·53156	·47825	·41008	·36824	·31946	·26179	·19244	·10729
20	0·58832	0·54574	0·49256	0·42409	0·38179	0·33221	0·27321	0·20170	0·11307
21	·60129	·55909	·50613	·43746	·39478	·34450	·28430	·21078	·11878
22	·61348	·57169	·51900	·45025	·40726	·35637	·29507	·21966	·12443
23	·62495	·58360	·53122	·46247	·41925	·36783	·30554	·22837	·13003
24	·63576	·59487	·54285	·47418	·43078	·37891	·31572	·23690	·13556
25	0·64597	0·60556	0·55392	0·48539	0·44186	0·38961	0·32561	0·24525	0·14103
26	·65563	·61570	·56447	·49615	·45253	·39996	·33524	·25343	·14645
27	·66478	·62533	·57455	·50647	·46281	·40997	·34460	·26145	·15180
28	·67346	·63450	·58417	·51639	·47272	·41967	·35371	·26931	·15710
29	·68170	·64323	·59337	·52591	·48228	·42906	·36259	·27701	·16234
30	0·68954	0·65156	0·60217	0·53508	0·49150	0·43815	0·37122	0·28455	0·16752
40	·75095	·71758	·67308	·61054	·56853	·51555	·44650	·35245	·21626
60	·82163	·79529	·75915	·70620	·66914	·62056	·55390	·45636	·29913
120	·90370	·88794	·86554	·83103	·80557	·77041	·71852	·63381	·46918
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

BETA DISTRIBUTION: 10 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9
1	0·024472	0·010000	0·0061812	0·0044577	0·0034818	0·0028553	0·0024193	0·0020986	0·0018528
2	·19000	·10000	·067830	·051317	·041268	·034511	·029654	·025996	·023141
3	·35136	·21544	·15648	·12310	·10154	·086434	·075257	·066647	·059809
4	·46812	·31623	·24136	·19580	·16493	·14256	·12558	·11224	·10147
5	0·55185	0·39811	0·31529	0·26204	0·22457	0·19664	0·17498	0·15766	0·14349
6	·61375	·46416	·37816	·32046	·27858	·24664	·22139	·20091	·18394
7	·66104	·51795	·43151	·37151	·32685	·29210	·26421	·24127	·22207
8	·69821	·56234	·47700	·41611	·36982	·33319	·30339	·27860	·25764
9	·72814	·59948	·51610	·45522	·40811	·37029	·33915	·31299	·29067
10	0·75273	0·63096	0·54996	0·48968	0·44232	0·40382	0·37178	0·34462	0·32128
11	·77328	·65793	·57954	·52022	·47300	·43419	·40159	·37374	·34963
12	·79069	·68129	·60555	·54744	·50062	·46178	·42889	·40058	·37592
13	·80564	·70170	·62860	·57181	·52560	·48693	·45393	·42535	·40032
14	·81861	·71969	·64915	·59375	·54827	·50992	·47697	·44827	·42299
15	0·82996	0·73564	0·66758	0·61360	0·56893	0·53100	0·49822	0·46951	0·44410
16	·83998	·74989	·68419	·63164	·58783	·55040	·51787	·48924	·46380
17	·84889	·76270	·69923	·64809	·60517	·56829	·53608	·50760	·48219
18	·85686	·77426	·71293	·66315	·62114	·58484	·55300	·52473	·49942
19	·86403	·78476	·72544	·67699	·63588	·60020	·56876	·54074	·51557
20	0·87052	0·79433	0·73691	0·68976	0·64954	0·61448	0·58347	0·55574	0·53073
21	·87643	·80309	·74747	·70156	·66222	·62779	·59722	·56980	·54500
22	·88181	·81113	·75722	·71250	·67403	·64022	·61011	·58302	·55845
23	·88675	·81855	·76625	·72268	·68504	·65187	·62222	·59547	·57115
24	·89129	·82540	·77464	·73216	·69535	·66279	·63361	·60721	·58314
25	0·89549	0·83176	0·78245	0·74103	0·70500	0·67305	0·64434	0·61829	0·59450
26	·89937	·83768	·78973	·74933	·71407	·68271	·65446	·62878	·60526
27	·90297	·84319	·79655	·75711	·72260	·69183	·66403	·63871	·61548
28	·90633	·84834	·80294	·76443	·73064	·70044	·67309	·64813	·62518
29	·90946	·85317	·80894	·77132	·73823	·70858	·68168	·65708	·63442
30	0·91239	0·85770	0·81459	0·77783	0·74541	0·71630	0·68984	0·66559	0·64322
40	·93381	·89125	·85693	·82706	·80025	·77578	·75319	·73219	·71255
60	·95555	·92612	·90182	·88023	·86048	·84212	·82490	·80864	·79321
120	·97761	·96235	·94944	·93773	·92679	·91643	·90653	·89702	·88785
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}10$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 10 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120
1	0·0016585	0·0013709	0·0010878	$0{\cdot}0^{3}80919$	$0{\cdot}0^{3}67157$	$0{\cdot}0^{3}53506$	$0{\cdot}0^{3}39965$	$0{\cdot}0^{3}26535$	$0{\cdot}0^{3}13213$
2	·020852	·017407	·013950	·010481	·0087416	·0069994	·0052542	·0035059	·0017545
3	·054245	·045740	·037035	·028119	·023579	·018982	·014327	·0096132	·0048379
4	·092595	·078823	·064482	·049452	·041691	·033749	·025617	·017288	·0087521
5	0·13167	0·11307	0·093336	0·072324	0·061295	0·049889	0·038083	0·025851	0·013167
6	·16964	·14685	·12228	·095653	·081477	·066668	·051174	·034941	·017906
7	·20573	·17941	·15059	·11886	·10173	·083668	·064573	·044345	·022866
8	·23966	·21040	·17792	·14161	·12177	·10064	·078083	·053928	·027978
9	·27139	·23970	·20411	·16374	·14141	·11743	·091577	·063600	·033196
10	0·30097	0·26732	0·22908	0·18513	0·16056	0·13394	0·10497	0·073298	0·038489
11	·32853	·29330	·25284	·20576	·17915	·15010	·11820	·082977	·043832
12	·35422	·31772	·27540	·22559	·19716	·16587	·13123	·092604	·049206
13	·37817	·34068	·29682	·24464	·21457	·18124	·14403	·10215	·054597
14	·40053	·36228	·31715	·26292	·23139	·19619	·15659	·11161	·059993
15	0·42143	0·38261	0·33645	0·28045	0·24762	0·21072	0·16889	0·12096	0·065386
16	·44100	·40176	·35478	·29726	·26327	·22483	·18093	·13019	·070768
17	·45934	·41983	·37219	·31338	·27837	·23853	·19270	·13930	·076134
18	·47657	·43689	·38875	·32885	·29293	·25182	·20420	·14828	·081478
19	·49277	·45302	·40451	·34369	·30697	·26471	·21544	·15712	·086796
20	0·50803	0·46829	0·41952	0·35793	0·32051	0·27721	0·22642	0·16583	0·092085
21	·52243	·48276	·43382	·37161	·33358	·28934	·23713	·17440	·097342
22	·53603	·49649	·44746	·38475	·34619	·30111	·24759	·18283	·10257
23	·54889	·50953	·46049	·39738	·35836	·31253	·25781	·19112	·10775
24	·56108	·52193	·47294	·40954	·37012	·32361	·26778	·19928	·11290
25	0·57263	0·53373	0·48485	0·42123	0·38147	0·33437	0·27751	0·20730	0·11801
26	·58361	·54498	·49624	·43248	·39245	·34481	·28701	·21518	·12308
27	·59405	·55571	·50716	·44333	·40306	·35495	·29629	·22293	·12811
28	·60398	·56595	·51763	·45378	·41332	·36479	·30534	·23054	·13310
29	·61344	·57574	·52767	·46386	·42325	·37436	·31419	·23803	·13804
30	0·62247	0·58511	0·53731	0·47359	0·43286	0·38366	0·32283	0·24539	0·14295
40	·69412	·66034	·61599	·55476	·51428	·46386	·39910	·31243	·18960
60	·77851	·75104	·71386	·66029	·62333	·57545	·51067	·41750	·27063
120	·87897	·86198	·83814	·80192	·77553	·73946	·68688	·60235	·44158
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

Beta Distribution: 5 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9
1	0·0061558	0·0025000	0·0015429	0·0011119	$0{\cdot}0^{3}86820$	$0{\cdot}0^{3}71179$	$0{\cdot}0^{3}60300$	$0{\cdot}0^{3}52300$	$0{\cdot}0^{3}46170$
2	·097500	·050000	·033617	·025321	·020308	·016952	·014548	·012741	·011334
3	·22852	·13572	·097308	·076010	·062412	·052962	·046007	·040671	·036447
4	·34163	·22361	·16825	·13535	·11338	·097611	·085728	·076440	·068979
5	0·43074	0·30171	0·23553	0·19403	0·16528	0·14408	0·12778	0·11482	0·10427
6	·50053	·36840	·29599	·24860	·21477	·18926	·16927	·15316	·13989
7	·55593	·42489	·34929	·29811	·26063	·23182	·20890	·19019	·17461
8	·60071	·47287	·39607	·34259	·30260	·27134	·24613	·22532	·20783
9	·63751	·51390	·43716	·38245	·34080	·30777	·28082	·25835	·23930
10	0·66824	0·54928	0·47338	0·41820	0·37553	0·34126	0·31301	0·28924	0·26894
11	·69425	·58003	·50546	·45033	·40712	·37203	·34283	·31807	·29677
12	·71654	·60696	·53402	·47930	·43590	·40031	·37044	·34494	·32286
13	·73583	·63073	·55958	·50551	·46219	·42635	·39604	·37000	·34732
14	·75268	·65184	·58256	·52932	·48626	·45036	·41980	·39338	·37025
15	0·76754	0·67070	0·60333	0·55102	0·50836	0·47255	0·44187	0·41521	0·39176
16	·78072	·68766	·62217	·57086	·52872	·49310	·46242	·43563	·41196
17	·79249	·70297	·63933	·58907	·54750	·51217	·48158	·45474	·43094
18	·80307	·71687	·65503	·60584	·56490	·52991	·49949	·47267	·44880
19	·81263	·72954	·66944	·62131	·58103	·54645	·51624	·48951	·46564
20	0·82131	0·74113	0·68271	0·63564	0·59605	0·56189	0·53194	0·50535	0·48152
21	·82923	·75178	·69496	·64894	·61004	·57635	·54669	·52027	·49652
22	·83647	·76160	·70632	·66132	·62312	·58990	·56056	·53434	·51071
23	·84313	·77067	·71687	·67287	·63536	·60263	·57363	·54764	·52415
24	·84927	·77908	·72669	·68366	·64684	·61461	·58596	·56022	·53689
25	0·85494	0·78690	0·73586	0·69377	0·65764	0·62590	0·59761	0·57213	0·54898
26	·86021	·79418	·74444	·70327	·66780	·63656	·60864	·58343	·56048
27	·86511	·80099	·75249	·71219	·67738	·64663	·61909	·59416	·57141
28	·86967	·80736	·76004	·72060	·68643	·65617	·62900	·60436	·58183
29	·87394	·81334	·76715	·72854	·69499	·66522	·63842	·61407	·59177
30	0·87794	0·81896	0·77386	0·73604	0·70311	0·67381	0·64738	0·62332	0·60125
40	·90734	·86089	·82447	·79327	·76559	·74053	·71758	·69636	·67663
60	·93748	·90497	·87881	·85591	·83517	·81606	·79824	·78150	·76569
120	·96837	·95130	·93720	·92458	·91290	·90192	·89148	·88150	·87191
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}05$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 5 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120
1	$0{\cdot}0^{3}41325$	$0{\cdot}0^{3}34155$	$0{\cdot}0^{3}27098$	$0{\cdot}0^{3}20156$	$0{\cdot}0^{3}16727$	$0{\cdot}0^{3}13326$	$0{\cdot}0^{4}99535$	$0{\cdot}0^{4}66082$	$0{\cdot}0^{4}32904$
2	·010206	·0085124	·0068158	·0051162	·0042653	·0034137	·0025614	·0017083	$\cdot0^{3}85452$
3	·033020	·027794	·022465	·017026	·014264	·011472	·0086511	·0057991	·0029157
4	·062850	·053375	·043541	·033319	·028053	·022679	·017191	·011585	·0058567
5	0·095510	0·081790	0·067312	0·051995	0·043994	0·035747	0·027240	0·018458	0·0093841
6	·12876	·11111	·092207	·071870	·061103	·049898	·038224	·026043	·013317
7	·16142	·14029	·11733	·092238	·078783	·064651	·049781	·034103	·017540
8	·19290	·16875	·14216	·11267	·096658	·079695	·061675	·042481	·021976
9	·22292	·19618	·16638	·13288	·11449	·094827	·073748	·051068	·026572
10	0·25137	0·22244	0·18984	0·15272	0·13211	0·10991	0·085885	0·059785	0·031288
11	·27823	·24746	·21244	·17207	·14943	·12484	·098008	·068575	·036094
12	·30354	·27125	·23413	·19086	·16636	·13955	·11006	·077394	·040967
13	·32737	·29383	·25492	·20908	·18288	·15401	·12199	·086209	·045889
14	·34981	·31524	·27481	·22669	·19895	·16818	·13377	·094994	·050847
15	0·37095	0·33554	0·29382	0·24370	0·21457	0·18203	0·14539	0·10373	0·055827
16	·39086	·35480	·31199	·26011	·22972	·19556	·15682	·11240	·060821
17	·40965	·37307	·32936	·27594	·24441	·20877	·16805	·12099	·065820
18	·42738	·39041	·34596	·29120	·25865	·22164	·17908	·12950	·070818
19	·44414	·40689	·36183	·30591	·27244	·23418	·18989	·13791	·075809
20	0·45999	0·42256	0·37701	0·32009	0·28580	0·24639	0·20050	0·14622	0·080789
21	·47501	·43746	·39154	·33375	·29874	·25828	·21088	·15442	·085753
22	·48925	·45165	·40544	·34693	·31126	·26985	·22106	·16252	·090698
23	·50276	·46518	·41877	·35964	·32340	·28112	·23102	·17051	·095621
24	·51560	·47808	·43154	·37190	·33515	·29208	·24078	·17838	·10052
25	0·52782	0·49040	0·44379	0·38373	0·34653	0·30275	0·25032	0·18615	0·10539
26	·53945	·50217	·45554	·39516	·35756	·31314	·25966	·19379	·11024
27	·55054	·51343	·46683	·40619	·36826	·32325	·26880	·20133	·11505
28	·56112	·52420	·47768	·41685	·37862	·33309	·27775	·20875	·11983
29	·57122	·53452	·48812	·42715	·38867	·34267	·28650	·21606	·12458
30	0·58088	0·54442	0·49816	0·43711	0·39842	0·35200	0·29507	0·22326	0·12930
40	·65819	·62460	·58083	·52099	·48175	·43321	·37136	·28936	·17453
60	·75070	·72282	·68535	·63185	·59522	·54807	·48477	·39458	·25416
120	·86266	·84504	·82047	·78342	·75661	·72016	·66738	·58326	·42519
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

Beta Distribution: 2·5 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	0·0015413	0·0^{3}62500	0·0^{3}38558	0·0^{3}27783	0·0^{3}21691	0·0^{3}17782	0·0^{3}15064	0·0^{3}13065	0·0^{3}11533
2	·049375	·025000	·016737	·012579	·010076	·0084038	·0072076	·0063095	·0056104
3	·14675	·085499	·060830	·047316	·038748	·032820	·028471	·025143	·022513
4	·24664	·15811	·11786	·094299	·078706	·067586	·059243	·052745	·047539
5	0·33318	0·22865	0·17674	0·14471	0·12275	0·10669	0·094390	0·084663	0·076770
6	·40505	·29240	·23259	·19412	·16695	·14663	·13081	·11812	·10770
7	·46442	·34855	·28375	·24063	·20942	·18562	·16681	·15153	·13886
8	·51378	·39764	·32993	·28358	·24933	·22278	·20151	·18405	·16944
9	·55524	·44054	·37137	·32290	·28642	·25774	·23450	·21523	·19897
10	0·59043	0·47818	0·40855	0·35877	0·32071	0·29042	0·26561	0·24486	0·22722
11	·62062	·51135	·44194	·39146	·35234	·32085	·29482	·27288	·25409
12	·64677	·54074	·47202	·42128	·38149	·34914	·32219	·29930	·27957
13	·66961	·56693	·49920	·44853	·40838	·37545	·34779	·32416	·30368
14	·68973	·59038	·52385	·47349	·43321	·39991	·37175	·34755	·32646
15	0·70756	0·61149	0·54628	0·49641	0·45618	0·42268	0·39418	0·36955	0·34799
16	·72349	·63058	·56676	·51750	·47746	·44390	·41520	·39026	·36833
17	·73778	·64792	·58553	·53697	·49723	·46372	·43490	·40976	·38756
18	·75069	·66373	·60278	·55498	·51561	·48224	·45341	·42814	·40575
19	·76239	·67821	·61869	·57169	·53276	·49959	·47081	·44549	·42297
20	0·77305	0·69150	0·63339	0·58722	0·54877	0·51586	0·48719	0·46187	0·43928
21	·78280	·70376	·64702	·60169	·56375	·53115	·50263	·47736	·45475
22	·79176	·71509	·65970	·61520	·57780	·54553	·51720	·49202	·46943
23	·80001	·72559	·67150	·62785	·59100	·55908	·53098	·50592	·48338
24	·80763	·73535	·68253	·63970	·60341	·57187	·54401	·51911	·49664
25	0·81469	0·74445	0·69285	0·65084	0·61511	0·58396	0·55636	0·53163	0·50927
26	·82126	·75295	·70253	·66132	·62615	·59540	·56808	·54354	·52130
27	·82738	·76090	·71162	·67119	·63658	·60624	·57922	·55488	·53278
28	·83310	·76836	·72018	·68052	·64646	·61652	·58980	·56568	·54373
29	·83845	·77538	·72825	·68933	·65582	·62630	·59988	·57599	·55420
30	0·84347	0·78198	0·73587	0·69768	0·66471	0·63559	0·60948	0·58582	0·56421
40	·88059	·83157	·79381	·76184	·73369	·70839	·68532	·66411	·64446
60	·91904	·88430	·85681	·83298	·81156	·79193	·77372	·75668	·74065
120	·95883	·94037	·92535	·91201	·89975	·88828	·87743	·86708	·85717
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}025$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 2·5 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	10	12	15	20	24	30	40	60	120
1	$0{\cdot}0^{3}10323$	$0{\cdot}0^{4}85313$	$0{\cdot}0^{4}67686$	$0{\cdot}0^{4}50345$	$0{\cdot}0^{4}41780$	$0{\cdot}0^{4}33285$	$0{\cdot}0^{4}24860$	$0{\cdot}0^{4}16505$	$0{\cdot}0^{5}82180$
2	·0050508	·0042107	·0033700	·0025286	·0021076	·0016864	·0012651	$\cdot 0^{3}84357$	$\cdot 0^{3}42187$
3	·020382	·017139	·013838	·010477	·0087725	·0070519	·0053148	·0035607	·0017893
4	·043272	·036693	·029885	·022831	·019207	·015514	·011749	·0079110	·0039956
5	0·070233	0·060028	0·049302	0·038002	0·032119	0·026068	0·019841	0·013428	0·0068184
6	·098988	·085233	·070563	·054861	·046579	·037985	·029056	·019767	·010092
7	·12818	·11113	·092695	·072663	·061969	·050772	·039029	·026691	·013702
8	·15701	·13700	·11508	·090920	·077871	·064092	·049508	·034033	·017569
9	·18504	·16240	·13732	·10931	·094004	·077712	·060314	·041675	·021634
10	0·21201	0·18709	0·15917	0·12760	0·11017	0·091466	0·071319	0·049528	0·025854
11	·23780	·21091	·18048	·14565	·12623	·10523	·082426	·057528	·030196
12	·26238	·23379	·20115	·16336	·14210	·11893	·093564	·065622	·034634
13	·28573	·25571	·22112	·18067	·15770	·13249	·10468	·073771	·039147
14	·30790	·27667	·24039	·19753	·17299	·14588	·11573	·081944	·043718
15	0·32893	0·29668	0·25893	0·21392	0·18793	0·15905	0·12669	0·090115	0·048335
16	·34888	·31578	·27676	·22983	·20252	·17198	·13753	·098266	·052985
17	·36779	·33400	·29389	·24525	·21674	·18466	·14823	·10638	·057659
18	·38574	·35138	·31034	·26019	·23058	·19708	·15878	·11444	·062348
19	·40278	·36797	·32614	·27465	·24404	·20922	·16916	·12244	·067047
20	0·41896	0·38380	0·34132	0·28864	0·25713	0·22110	0·17938	0·13038	0·071749
21	·43435	·39893	·35589	·30218	·26985	·23270	·18943	·13823	·076450
22	·44900	·41338	·36990	·31528	·28221	·24402	·19930	·14601	·081144
23	·46294	·42720	·38335	·32795	·29422	·25508	·20899	·15370	·085828
24	·47623	·44042	·39629	·34021	·30588	·26587	·21850	·16130	·090500
25	0·48891	0·45307	0·40874	0·35207	0·31721	0·27640	0·22783	0·16881	0·095156
26	·50101	·46520	·42071	·36355	·32821	·28667	·23698	·17622	·099794
27	·51257	·47682	·43223	·37466	·33890	·29669	·24596	·18354	·10441
28	·52363	·48797	·44334	·38542	·34928	·30647	·25476	·19076	·10901
29	·53421	·49867	·45403	·39584	·35937	·31601	·26339	·19789	·11358
30	0·54435	0·50895	0·46434	0·40594	0·36918	0·32532	0·27185	0·20492	0·11812
40	·62616	·59296	·54999	·49168	·45370	·40697	·34780	·26997	·16201
60	·72550	·69743	·65992	·60674	·57056	·52422	·46239	·37498	·24027
120	·84764	·82954	·80442	·76678	·73968	·70299	·65017	·56658	·41107
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

Beta Distribution: 1 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	0·0^{3}24672	0·0^{3}10000	0·0^{4}61686	0·0^{4}44446	0·0^{4}34699	0·0^{4}28446	0·0^{4}24097	0·0^{4}20897	0·0^{4}18449
2	·019900	·010000	·0066778	·0050126	·0040121	·0033445	·0028674	·0025094	·0022309
3	·080827	·046416	·032834	·025458	·020807	·017599	·015252	·013458	·012043
4	·15875	·10000	·073960	·058903	·049014	·041999	·036754	·032682	·029426
5	0·23520	0·15849	0·12142	0·098877	0·083563	0·072429	0·063948	0·057264	0·051857
6	·30387	·21544	·16079	·14087	·12065	·10564	·094014	·084730	·077136
7	·36370	·26827	·21636	·18236	·15801	·13959	·12511	·11341	·10375
8	·41540	·31623	·25997	·22207	·19437	·17307	·15612	·14227	·13073
9	·46009	·35938	·30024	·25945	·22910	·20543	·18637	·17066	·15745
10	0·49889	0·39811	0·33719	0·29431	0·26191	0·23632	0·21551	0·19820	0·18355
11	·53279	·43288	·37099	·32667	·29271	·26560	·24335	·22469	·20879
12	·56258	·46416	·40191	·35664	·32153	·29323	·26981	·25003	·23307
13	·58893	·49239	·43020	·38437	·34845	·31924	·29487	·27417	·25631
14	·61238	·51795	·45615	·41006	·37358	·34369	·31858	·29712	·27851
15	0·63336	0·54117	0·47999	0·43387	0·39706	0·36666	0·34098	0·31891	0·29968
16	·65224	·56234	·50194	·45597	·41899	·38826	·36214	·33958	·31985
17	·66930	·58171	·52219	·47651	·43951	·40857	·38213	·35920	·33905
18	·68479	·59948	·54094	·49565	·45872	·42768	·40103	·37781	·35733
19	·69892	·61585	·55832	·51350	·47674	·44568	·41890	·39547	·37474
20	0·71185	0·63096	0·57447	0·53018	0·49366	0·46266	0·43581	0·41224	0·39131
21	·72372	·64495	·58952	·54581	·50958	·47868	·45184	·42818	·40711
22	·73467	·65793	·60357	·56046	·52456	·49383	·46703	·44333	·42217
23	·74479	·67002	·61671	·57422	·53869	·50816	·48144	·45775	·43653
24	·75417	·68129	·62903	·58717	·55204	·52174	·49514	·47149	·45025
25	0·76290	0·69183	0·64059	0·59938	0·56466	0·53461	0·50816	0·48458	0·46335
26	·77103	·70170	·65147	·61090	·57660	·54683	·52055	·49706	·47587
27	·77862	·71097	·66172	·62180	·58793	·55845	·53236	·50899	·48785
28	·78573	·71969	·67139	·63211	·59868	·56951	·54362	·52038	·49932
29	·79240	·72790	·68054	·64188	·60890	·58004	·55437	·53127	·51031
30	0·79867	0·73564	0·68919	0·65116	0·61862	0·59008	0·56464	0·54170	0·52085
40	·84541	·79433	·75561	·72316	·69482	·66950	·64656	·62555	·60617
60	·89449	·85770	·82898	·80433	·78233	·76227	·74376	·72651	·71034
120	·94599	·92612	·91014	·89607	·88321	·87124	·85995	·84924	·83900
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}01$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 1 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	10	12	15	20	24	30	40	60	120
1	$0{\cdot}0^{4}16513$	$0{\cdot}0^{4}13647$	$0{\cdot}0^{4}10827$	$0{\cdot}0^{5}80531$	$0{\cdot}0^{5}66831$	$0{\cdot}0^{5}53242$	$0{\cdot}0^{5}39766$	$0{\cdot}0^{5}26400$	$0{\cdot}0^{5}13145$
2	·0020080	·0016737	·0013391	·0010045	$\cdot 0^{3}83718$	$\cdot 0^{3}66980$	$\cdot 0^{3}50239$	$\cdot 0^{3}33496$	$\cdot 0^{3}16749$
3	·010898	·0091569	·0073877	·0055887	·0046777	·0037588	·0028317	·0018964	$\cdot 0^{3}95252$
4	·026763	·022665	·018435	·014065	·011824	·0095436	·0072226	·0048595	·0024525
5	0·047389	0·040434	0·033149	0·025503	0·021534	0·017459	0·013275	0·0089747	0·0045520
6	·070804	·060840	·050258	·038982	·033057	·026923	·020567	·013973	·0071235
7	·095627	·082714	·068820	·053801	·045816	·037481	·028767	·019640	·010065
8	·12095	·10526	·088177	·069456	·059390	·048797	·037625	·025815	·013300
9	·14619	·12796	·10787	·085584	·073472	·060623	·046956	·032376	·016768
10	0·17097	0·15044	0·12760	0·10193	0·087838	0·072776	0·056621	0·039229	0·020426
11	·19506	·17250	·14713	·11830	·10232	·085117	·066512	·046303	·024237
12	·21834	·19398	·16633	·13458	·11681	·097542	·076547	·053541	·028173
13	·24073	·21479	·18511	·15065	·13120	·10997	·086660	·060897	·032212
14	·26220	·23489	·20338	·16646	·14544	·12235	·096802	·068334	·036335
15	0·28276	0·25426	0·22113	0·18196	0·15948	0·13462	0·10693	0·075824	0·040526
16	·30240	·27289	·23833	·19711	·17327	·14676	·11702	·083341	·044772
17	·32117	·29079	·25497	·21189	·18681	·15873	·12704	·000866	·049062
18	·33910	·30797	·27105	·22630	·20005	·17053	·13697	·098383	·053386
19	·35622	·32446	·28658	·24032	·21301	·18212	·14680	·10588	·057738
20	0·37257	0·34029	0·30157	0·25395	0·22567	0·19351	0·15651	0·11334	0·062109
21	·38818	·35548	·31603	·26721	·23803	·20468	·16609	·12076	·066494
22	·40311	·37005	·32909	·28008	·25008	·21563	·17554	·12812	·070888
23	·41738	·38405	·34345	·29258	·26184	·22636	·18486	·13543	·075285
24	·43103	·39749	·35645	·30472	·27329	·23687	·19403	·14268	·079683
25	0·44410	0·41040	0·36899	0·31651	0·28446	0·24716	0·20305	0·14986	0·084077
26	·45661	·42280	·38109	·32795	·29534	·25722	·21193	·15697	·088465
27	·46861	·43473	·39278	·33906	·30594	·26707	·22066	·16401	·092843
28	·48011	·44621	·40407	·34985	·31626	·27670	·22925	·17096	·097210
29	·49115	·45726	·41497	·36032	·32632	·28612	·23768	·17785	·10156
30	0·50175	0·46789	0·42552	0·37049	0·33612	0·29534	0·24597	0·18465	0·10590
40	·58819	·55573	·51398	·45778	·42144	·37700	·32111	·24819	·14811
60	·69511	·66701	·62969	·57717	·54167	·49647	·43655	·35258	·22459
120	·82918	·81062	·78497	·74677	·71942	·68259	·62988	·54709	·39479
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

Beta Distribution: 0·5 per cent Points for x

$\nu_1 = 2q \qquad \nu_2 = 2p$

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9
1	$0{\cdot}0^{4}61684$	$0{\cdot}0^{4}25000$	$0{\cdot}0^{4}15421$	$0{\cdot}0^{4}11111$	$0{\cdot}0^{5}86745$	$0{\cdot}0^{5}71112$	$0{\cdot}0^{5}60240$	$0{\cdot}0^{5}52245$	$0{\cdot}0^{5}46121$
2	·0099750	·0050000	·0033361	·0025031	·0020030	·0016695	·0014311	·0012524	·0011133
3	·051237	·029240	·020632	·015976	·013046	·011028	·0095530	·0084269	·0075388
4	·11321	·070711	·052099	·041400	·034399	·029445	·025748	·022881	·020592
5	0·17996	0·12011	0·091593	0·074378	0·062737	0·054301	0·047891	0·042849	0·038776
6	·24356	·17100	·13408	·11088	·094759	·082829	·073619	·066279	·060287
7	·30126	·22007	·17656	·14830	·12818	·11303	·10116	·091593	·083707
8	·35261	·26591	·21745	·18510	·16159	·14360	·12933	·11770	·10804
9	·39799	·30808	·25604	·22046	·19415	·17373	·15736	·14389	·13261
10	0·43809	0·34657	0·29204	0·25399	0·22542	0·20297	0·18478	0·16970	0·15697
11	·47360	·38162	·32543	·28554	·25517	·23105	·21132	·19484	·18083
12	·50517	·41352	·35632	·31509	·28332	·25783	·23682	·21914	·20402
13	·53337	·44258	·38487	·34270	·30986	·28328	·26120	·24250	·22642
14	·55865	·46912	·41127	·36848	·33484	·30739	·28444	·26489	·24798
15	0·58144	0·49340	0·43569	0·39255	0·35833	0·33022	0·30656	0·28629	0·26869
16	·60206	·51567	·45832	·41503	·38042	·35180	·32757	·30672	·28852
17	·62080	·53616	·47932	·43605	·40120	·37221	·34754	·32620	·30752
18	·63789	·55505	·49885	·45571	·42076	·39151	·36650	·34478	·32568
19	·65354	·57251	·51703	·47414	·43917	·40976	·38451	·36248	·34305
20	0·66792	0·58870	0·53400	0·49144	0·45654	0·42705	0·40162	0·37936	0·35966
21	·68117	·60375	·54986	·50768	·47293	·44343	·41789	·39546	·37554
22	·69341	·61775	·56472	·52297	·48841	·45896	·43336	·41082	·39073
23	·70477	·63083	·57866	·53738	·50306	·47369	·44810	·42547	·40527
24	·71532	·64305	·59176	·55098	·51692	·48769	·46213	·43947	·41918
25	0·72516	0·65451	0·60409	0·56382	0·53007	0·50100	0·47551	0·45285	0·43251
26	·73434	·66527	·61571	·57597	·54265	·51367	·48827	·46564	·44528
27	·74294	·67539	·62669	·58749	·55441	·52574	·50046	·47788	·45752
28	·75100	·68492	·63707	·59841	·56568	·53724	·51210	·48960	·46927
29	·75857	·69392	·64690	·60878	·57642	·54822	·52324	·50083	·48054
30	0·76570	0·70242	0·65622	0·61864	0·58665	0·55871	0·53389	0·51159	0·49137
40	·81920	·76727	·72823	·69571	·66744	·64229	·61956	·59882	·57973
60	·87598	·83811	·80877	·78370	·76142	·74119	·72256	·70526	·68907
120	·93619	·91548	·89893	·88442	·87120	·85892	·84739	·83645	·82602
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

This table gives the values of x for which $I_x(p, q) = 0{\cdot}005$ where $p = \frac{1}{2}\nu_2$, $q = \frac{1}{2}\nu_1$.

BETA DISTRIBUTION: 0·5 PER CENT POINTS FOR x

$\nu_1 = 2q \qquad \nu_2 = 2p$

ν_2 \ ν_1	10	12	15	20	24	30	40	60	120
1	0·0^{5}41280	0·0^{5}34116	0·0^{5}27067	0·0^{5}20132	0·0^{5}16707	0·0^{5}13310	0·0^{6}99411	0·0^{6}65998	0·0^{6}32862
2	·0010020	·0^{3}83507	·0^{3}66812	·0^{3}50113	·0^{3}41762	·0^{3}33411	·0^{3}25060	·0^{3}16707	·0^{4}83539
3	·0068204	·0057290	·0046206	·0034943	·0029242	·0023493	·0017696	·0011848	·0^{3}59503
4	·018721	·015844	·012879	·0098197	·0082522	·0066584	·0050373	·0033880	·0017093
5	0·035415	0·030191	0·024729	0·019006	0·016040	0·012998	0·0098777	0·0066743	0·0033833
6	·055299	·047464	·039162	·030337	·025709	·020924	·015973	·010844	·0055240
7	·077090	·066593	·055329	·043189	·036749	·030038	·023034	·015712	·0080441
8	·099867	·086787	·072586	·057076	·048760	·040023	·030828	·021129	·010873
9	·12300	·10749	·090464	·071635	·061433	·050635	·039174	·026976	·013953
10	0·14606	0·12831	0·10862	0·086595	0·074540	0·061684	0·047930	0·033162	0·017241
11	·16876	·14898	·12683	·10175	·087903	·073027	·056985	·039611	·020700
12	·19092	·16931	·14489	·11696	·10139	·084550	·066252	·046265	·024302
13	·21242	·18919	·16270	·13210	·11490	·096167	·075662	·053076	·028022
14	·23320	·20853	·18017	·14710	·12835	·10781	·085158	·060005	·031841
15	0·25323	0·22728	0·19725	0·16190	0·14170	0·11942	0·094697	0·067019	0·035743
16	·27248	·24543	·21388	·17644	·15488	·13097	·10424	·074093	·039714
17	·29098	·26295	·23006	·19070	·16787	·14241	·11377	·081204	·043741
18	·30872	·27986	·24576	·20465	·18065	·15373	·12324	·088334	·047815
19	·32574	·29615	·26099	·21829	·19319	·16489	·13265	·095467	·051928
20	0·34206	0·31184	0·27575	0·23160	0·20549	0·17590	0·14198	0·10259	0·056070
21	·35770	·32696	·29004	·24458	·21753	·18673	·15122	·10969	·060237
22	·37269	·34151	·30387	·25723	·22932	·19738	·16036	·11676	·064421
23	·38707	·35552	·31725	·26954	·24084	·20784	·16938	·12380	·068619
24	·40087	·36901	·33020	·28153	·25210	·21811	·17829	·13078	·072825
25	0·41411	0·38200	0·34272	0·29320	0·26309	0·22818	0·18707	0·13772	0·077036
26	·42682	·39452	·35484	·30456	·27383	·23806	·19573	·14461	·081248
27	·43903	·40658	·36657	·31560	·28432	·24775	·20426	·15143	·085459
28	·45076	·41821	·37792	·32635	·29455	·25724	·21266	·15819	·089665
29	·46204	·42942	·38891	·33681	·30454	·26654	·22093	·16489	·093863
30	0·47289	0·44024	0·39954	0·34698	0·31429	0·27565	0·22907	0·17152	0·098050
40	·56205	·53024	·48950	·43493	·39980	·35700	·30341	·23388	·13907
60	·67384	·64584	·60879	·55688	·52194	·47762	·41913	·33759	·21421
120	·81604	·79720	·77125	·73277	·70531	·66845	·61590	·53378	·38380
∞	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000	1·00000

For $\nu_1 = \infty$, $x = 0$

TABLE OF LAGRANGIAN COEFFICIENTS FOR HARMONIC INTERPOLATION IN CERTAIN TABLES OF PERCENTAGE POINTS

PREPARED BY L. J. COMRIE AND H. O. HARTLEY

This table was prepared to facilitate interpolation in the tables of percentage points of the Incomplete Beta-Function in the preceding paper. As, however, the main part of the table may be applied to interpolation in other tables with a similar lay-out, it is published separately.

The table is based on harmonic interpolation—a device introduced by R. A. Fisher in his tables of percentage points of the distribution of z, whose two parameters (degrees of freedom) n_1 and n_2 range from 1 to ∞. It consists in using values of n_1 and n_2 in harmonic progression, so that, with $1/n_1$ and $1/n_2$ as variables, z is tabulated at equidistant intervals near $1/n_1 = 0$ and also near $1/n_2 = 0$. This transformation renders the z-table (apart from its singularity at $n_1 = \infty$, $n_2 = \infty$) interpolable. As the percentage points of the Incomplete Beta-Function show a similar behaviour,* the values of ν_1 and ν_2 near the margin of the tables have been chosen in harmonic progression. Harmonic interpolation is, in fact, applicable to any table of percentage points (depending on a parameter n with an infinite range) in which the statistic can be adequately represented as a polynomial in $1/n$.† It is

Table of Lagrangian Coefficients

Column headings are the arguments of tabular values and row headings the arguments of the interpolate.

			Ordinary				
	7	8	9	10	12	15	20
	−	+	−	+	+	−	+
11	0·069 231	0·428 571	1·090 909	1·440 000	0·300 000	0·008 571	0·000 140
	8	9	10	12	15	20	24
	−	+	−	+	+	−	+
13	0·171 875	0·777 778	1·100 000	1·336 805	0·162 963	0·006 250	0·000 579
14	0·223 214	0·969 697	1·285 714	1·041 667	0·507 936	0·011 364	0·000 992
	9	10	12	15	20	24	30
	−	+	−	+	+	−	+
16	0·172 391	0·448 000	0·604 938	1·238 914	0·106 909	0·017 284	0·000 790
17	0·306 397	0·780 000	0·983 025	1·258 272	0·289 546	0·040 124	0·001 728
18	0·332 468	0·833 143	1·000 000	1·024 000	0·530 182	0·057 143	0·002 286
19	0·222 222	0·550 000	0·636 574	0·570 370	0·787 500	0·050 926	0·001 852
				Harmonic			
	10	12	15	20	24	30	40
	+	−	+	+	−	+	−
16	0·003 052	0·039 551	0·692 139	0·769 043	0·632 813	0·247 192	0·039 062
17	0·003 097	0·037 459	0·409 711	1·213 958	0·856 212	0·315 162	0·048 257
18	0·001 996	0·022 993	0·201 189	1·341 259	0·735 777	0·251 486	0·037 160
19	0·000 818	0·009 091	0·069 601	1·237 345	0·407 263	0·126 547	0·017 957

* This issue, pp. 168–81.

† It can be shown that any so-called "studentized" statistic has this property.

Table of Lagrangian Coefficients (continued)

	Harmonic						
	12	15	20	24	30	40	60
	+	−	+	+	−	+	−
21	0·000 600	0·010 497	0·629 840	0·537 463	0·209 947	0·060 616	0·008 075
22	0·000 579	0·009 653	0·337 838	0·864 864	0·253 378	0·068 640	0·008 890
23	0·000 306	0·004 908	0·130 868	1·005 068	0·168 259	0·042 230	0·005 305
	15	20	24	30	40	60	120
	+	−	+	+	−	+	−
25	0·000 730	0·040 858	0·817 152	0·306 432	0·108 954	0·029 184	0·003 686
26	0·000 993	0·050 984	0·611 813	0·573 575	0·174 804	0·044 986	0·005 570
27	0·000 927	0·044 488	0·415 220	0·778 537	0·191 640	0·047 184	0·005 740
28	0·000 670	0·030 498	0·243 980	0·914 925	0·162 653	0·038 122	0·004 546
29	0·000 335	0·014 607	0·105 172	0·985 984	0·095 611	0·021 204	0·002 477
	20 (10)	24 (12)	30 (15)	40 (20)	60 (30)	120 (60)	∞ (∞)
	+	−	+	+	−	+	−
31 (15·5)	0·003 664	0·041 459	0·906 909	0·179 142	0·062 545	0·016 304	0·002 015
32 (16·0)	0·005 875	0·063 446	0·793 076	0·352 478	0·113 297	0·028 839	0·003 525
33 (16·5)	0·006 875	0·071 503	0·670 341	0·510 736	0·148 964	0·036 984	0·004 469
34 (17·0)	0·006 948	0·070 033	0·547 129	0·648 450	0·168 348	0·040 717	0·004 863
35 (17·5)	0·006 358	0·062 423	0·429 158	0·762 947	0·171 663	0·040 391	0·004 768
36 (18·0)	0·005 335	0·051 212	0·320 073	0·853 529	0·160 037	0·036 580	0·004 268
37 (18·5)	0·004 062	0·038 250	0·221 984	0·920 823	0·135 121	0·029 955	0·003 453
38 (19·0)	0·002 684	0·024 848	0·135 887	0·966 308	0·098 827	0·021 212	0·002 416
39 (19·5)	0·001 305	0·011 904	0·061 908	0·091 960	0·053 141	0·011 022	0·001 240
	−	+	−	+	+	−	+
41 (20·5)	0·001 182	0·010 511	0·050 764	0·992 724	0·058 780	0·011 310	0·001 241
42 (21·0)	0·002 210	0·019 448	0·091 161	0·972 384	0·121 548	0·022 440	0·002 431
43 (21·5)	0·003 069	0·026 748	0·122 164	0·941 117	0·186 839	0·033 000	0·003 520
44 (22·0)	0·003 754	0·032 432	0·144 787	0·900 900	0·253 378	0·042 674	0·004 505
45 (22·5)	0·004 268	0·036 580	0·160 037	0·853 529	0·320 073	0·051 212	0·005 335
46 (23·0)	0·004 619	0·039 302	0·168 878	0·800 606	0·386 006	0·058 422	0·006 005
47 (23·5)	0·004 819	0·040 733	0·172 219	0·743 549	0·450 419	0·064 169	0·006 506
48 (24·0)	0·004 883	0·041 016	0·170 899	0·683 594	0·512 695	0·068 350	0·006 836
49 (24·5)	0·004 824	0·040 293	0·165 679	0·621 808	0·572 346	0·070 939	0·006 995
50 (25·0)	0·004 659	0·038 707	0·157 248	0·559 104	0·628 992	0·071 885	0·006 989
51 (25·5)	0·004 402	0·036 392	0·146 218	0·496 255	0·682 351	0·071 202	0·006 824
52 (26·0)	0·004 068	0·033 473	0·133 132	0·433 910	0·732 223	0·068 915	0·006 509
53 (26·5)	0·003 669	0·030 065	0·118 464	0·372 604	0·778 476	0·065 067	0·006 055
54 (27·0)	0·003 220	0·026 273	0·102 630	0·312 777	0·821 038	0·059 712	0·005 474
55 (27·5)	0·002 730	0·022 191	0·085 989	0·254 781	0·859 886	0·052 016	0·004 777
56 (28·0)	0·002 210	0·017 901	0·068 849	0·198 897	0·895 035	0·044 752	0·003 978
57 (28·5)	0·001 669	0·013 477	0·051 474	0·145 339	0·926 536	0·035 297	0·003 088
58 (29·0)	0·001 116	0·008 983	0·034 088	0·094 268	0·954 463	0·024 631	0·002 121
59 (29·5)	0·000 558	0·004 475	0·016 878	0·045 797	0·978 914	0·012 838	0·001 088
	+	−	+	−	+	+	−
61 (30·5)	0·000 552	0·004 402	0·016 417	0·043 083	1·017 846	0·013 801	0·001 131
62 (31·0)	0·001 093	0·008 695	0·032 268	0·083 441	1·032 585	0·028 485	0·002 295
63 (31·5)	0·001 619	0·012 854	0·047 471	0·121 085	1·044 357	0·043 973	0·003 481
64 (32·0)	0·002 128	0·016 853	0·061 959	0·156 045	1·053 303	0·060 189	0·004 681
65 (32·5)	0·002 616	0·020 675	0·075 683	0·188 368	1·059 570	0·077 060	0·005 886
66 (33·0)	0·003 082	0·024 304	0·088 608	0·218 113	1·063 300	0·094 516	0·007 089
67 (33·5)	0·003 524	0·027 730	0·100 709	0·245 348	1·064 636	0·112 490	0·008 281
68 (34·0)	0·003 940	0·030 045	0·111 971	0·270 151	1·063 721	0·130 919	0·009 455
69 (34·5)	0·004 329	0·033 942	0·122 388	0·292 605	1·060 692	0·149 745	0·010 607

	Harmonic						
	20 (10) +	24 (12) −	30 (15) +	40 (20) −	60 (30) +	120 (60) +	∞ (∞) −
70 (35·0)	0·004 692	0·036 719	0·131 960	0·312 795	1·055 683	0·168 909	0·011 730
71 (35·5)	0·005 027	0·039 275	0·140 696	0·330 812	1·048 823	0·188 360	0·012 819
72 (36·0)	0·005 335	0·041 610	0·148 605	0·346 746	1·040 238	0·208 048	0·013 870
73 (36·5)	0·005 615	0·043 725	0·155 705	0·360 690	1·030 048	0·227 925	0·014 878
74 (37·0)	0·005 867	0·045 623	0·162 013	0·372 737	1·018 370	0·247 951	0·015 841
75 (37·5)	0·006 093	0·047 309	0·167 552	0·382 976	1·005 312	0·268 083	0·016 755
76 (38·0)	0·006 292	0·048 787	0·172 345	0·391 499	0·990 981	0·288 285	0·017 617
77 (38·5)	0·006 465	0·050 062	0·176 416	0·398 393	0·975 477	0·308 523	0·018 426
78 (39·0)	0·006 613	0·051 141	0·179 793	0·403 745	0·958 894	0·328 764	0·019 178
79 (39·5)	0·006 736	0·052 030	0·182 502	0·407 639	0·941 324	0·348 979	0·019 872
80 (40·0)	0·006 836	0·052 734	0·184 570	0·410 156	0·922 851	0·369 141	0·020 508
81 (40·5)	0·006 912	0·053 262	0·186 027	0·411 376	0·903 557	0·389 225	0·021 083
82 (41·0)	0·006 967	0·053 621	0·186 898	0·411 373	0·883 518	0·409 208	0·021 597
83 (41·5)	0·007 000	0·053 816	0·187 212	0·410 223	0·862 805	0·429 071	0·022 049
84 (42·0)	0·007 012	0·053 855	0·186 997	0·407 993	0·841 486	0·448 793	0·022 440
85 (42·5)	0·007 005	0·053 746	0·186 279	0·404 753	0·819 625	0·468 357	0·022 767
86 (43·0)	0·006 980	0·053 495	0·185 083	0·400 567	0·797 283	0·487 749	0·023 033
87 (43·5)	0·006 936	0·053 110	0·183 437	0·395 496	0·774 514	0·506 954	0·023 235
88 (44·0)	0·006 875	0·052 596	0·181 366	0·389 600	0·751 371	0·525 960	0·023 376
89 (44·5)	0·006 798	0·051 961	0·178 892	0·382 934	0·727 905	0·544 755	0·023 455
90 (45·0)	0·006 706	0·051 212	0·176 040	0·375 552	0·704 161	0·563 329	0·023 472
91 (45·5)	0·006 600	0·050 354	0·172 833	0·367 506	0·680 182	0·581 673	0·023 428
92 (46·0)	0·006 479	0·049 393	0·169 293	0·358 843	0·656 009	0·599 780	0·023 325
93 (46·5)	0·006 346	0·048 337	0·165 440	0·349 609	0·631 680	0·617 642	0·023 162
94 (47·0)	0·006 200	0·047 190	0·161 295	0·339 848	0·607 229	0·635 254	0·022 940
95 (47·5)	0·006 043	0·045 959	0·156 878	0·329 602	0·582 689	0·652 611	0·022 660
96 (48·0)	0·005 875	0·044 647	0·152 207	0·318 909	0·558 090	0·669 708	0·022 324
97 (48·5)	0·005 696	0·043 262	0·147 299	0·307 806	0·533 462	0·686 542	0·021 931
98 (49·0)	0·005 509	0·041 806	0·142 173	0·296 330	0·508 829	0·703 109	0·021 484
99 (49·5)	0·005 312	0·040 287	0·136 844	0·284 511	0·484 217	0·719 408	0·020 983
100 (50·0)	0·005 107	0·038 707	0·131 328	0·272 384	0·459 648	0·735 437	0·020 429
101 (50·5)	0·004 894	0·037 072	0·125 640	0·259 976	0·435 143	0·751 194	0·019 823
102 (51·0)	0·004 675	0·035 385	0·119 793	0·247 316	0·410 721	0·766 679	0·019 167
103 (51·5)	0·004 448	0·033 651	0·113 802	0·234 429	0·386 400	0·781 891	0·018 461
104 (52·0)	0·004 216	0·031 873	0·107 680	0·221 342	0·362 196	0·796 831	0·017 708
105 (52·5)	0·003 978	0·030 056	0·101 437	0·208 077	0·338 125	0·811 499	0·016 906
106 (53·0)	0·003 735	0·028 201	0·095 087	0·194 656	0·314 199	0·825 895	0·016 059
107 (53·5)	0·003 486	0·026 314	0·088 639	0·181 099	0·290 433	0·840 022	0·015 167
108 (54·0)	0·003 234	0·024 397	0·082 104	0·167 427	0·266 837	0·853 880	0·014 231
109 (54·5)	0·002 978	0·022 452	0·075 492	0·153 658	0·243 423	0·867 470	0·013 253
110 (55·0)	0·002 719	0·020 484	0·068 812	0·139 809	0·220 199	0·880 796	0·012 233
111 (55·5)	0·002 456	0·018 494	0·062 074	0·125 896	0·197 175	0·893 858	0·011 173
112 (56·0)	0·002 190	0·016 485	0·055 284	0·111 933	0·174 358	0·906 660	0·010 074
113 (56·5)	0·001 922	0·014 459	0·048 452	0·097 936	0·151 755	0·919 203	0·008 937
114 (57·0)	0·001 651	0·012 420	0·041 584	0·083 918	0·129 374	0·931 491	0·007 762
115 (57·5)	0·001 379	0·010 368	0·034 688	0·069 891	0·107 219	0·943 525	0·006 552
116 (58·0)	0·001 106	0·008 307	0·027 770	0·055 866	0·085 295	0·955 309	0·005 307
177 (58·5)	0·000 831	0·006 238	0·020 837	0·041 854	0·063 608	0·966 845	0·004 029
118 (59·0)	0·000 554	0·004 162	0·013 894	0·027 867	0·042 161	0·978 137	0·002 717
119 (59·5)	0·000 278	0·002 083	0·006 947	0·013 913	0·020 957	0·989 188	0·001 374

particularly convenient if the high-order terms are so small that linear interpolation suffices. Unfortunately, however, if the interpolate is required to the same accuracy as the tabular values, linear interpolation in the above-mentioned tables is inadequate; hence this table has been prepared as the simplest method of preserving tabular accuracy in the interpolates. The interpolate is the sum of seven products of which the seven (Lagrangian) multipliers are taken from this table whilst the multiplicands are tabular entries in the table of percentage points. The examples below illustrate the use of the table.

The calculation of the Lagrangian coefficients follows the standard formulæ for interpolation by a polynomial of the sixth degree. Where interpolation is harmonic the coefficients are those for polynomials of the sixth degree in the reciprocal of the parameters used as argument.

In the early part of the table, however, ordinary Lagrangian coefficients are given, since the polynomial in the parameter itself is preferable in this range. The part of the table with row headings less than 30 has been specifically designed to meet the requirements of the tables of percentage points of the Incomplete Beta Function. In particular it will be noted that there are two rows for 16, 17, 18 and 19, one giving harmonic and the other ordinary Lagrangian coefficients; the application of both rows affords a good check, as will be seen from *Example* 2 on p. 162 in the preceding paper.

The table may be used not only for the progression 10, 12, 15, 20, 24, 30, 40, 60, 120, ∞, but also for submultiple progressions. The most important of these, namely 10, 12, 15, 20, 30, 60, ∞, is obtained by halving the last seven terms, and is catered for by the auxiliary arguments in brackets. Division by 4 yields the progression 5, 6, 7·5, 10, 15, 30, ∞, while division by 5 yields 3, 4, 4·8, 6, 8, 12, 24, ∞, from which we can select the first seven or the last seven values.

The missing values 7·5 or 4·8 can be found by ordinary interpolation from values in their immediate neighbourhood. If linear interpolation does not suffice we may use

$$f(7{\cdot}5) = \tfrac{1}{16}\{-f(6)+9f(7)+9f(8)-f(9)\},$$

which takes third differences into account, and

$$f(4{\cdot}8) = 0{\cdot}12f(4)+0{\cdot}96f(5)-0{\cdot}08f(6),$$

which takes second differences into account. Since $f(7{\cdot}5)$ and $f(4{\cdot}8)$ are not required to full tabular accuracy these formulæ will, as a rule, suffice.

Example 1. Find the 0·5 % point of the Incomplete Beta Function corresponding to $\nu_1 = 4$ and $\nu_2 = 96$.

In the accompanying table enter row 96, which gives the Lagrangian multipliers. The corresponding multiplicands are taken from the column of the table of 0·5 % points on p. 180 and are the entries for $\nu_2 = 20, 24, 30, 40, 60, 120$ and ∞, which correspond to the column headings in the Lagrangian table. The sign of each product is also given at the top of the columns. We have, therefore, the scheme shown alongside. The result is two units greater in the fifth decimal than the exact value obtained by inverse interpolation in Pearson's tables.

$$\begin{aligned} x(96, 4) = &+0{\cdot}491\,44 \times 0{\cdot}005\,875 \\ &-0{\cdot}550\,98 \times 0{\cdot}044\,647 \\ &+0{\cdot}618\,64 \times 0{\cdot}152\,207 \\ &-0{\cdot}695\,71 \times 0{\cdot}318\,909 \\ &+0{\cdot}783\,70 \times 0{\cdot}558\,090 \\ &+0{\cdot}884\,42 \times 0{\cdot}669\,708 \\ &-1{\cdot}000\,00 \times 0{\cdot}022\,324 \\ = &\ 0{\cdot}857\,94 \end{aligned}$$

Example 2. In Fisher's table of 1 % points of the distribution of z find the point corresponding to $n_1 = 12$ and $n_2 = 54$.

In Fisher's table we have the harmonic progression 10, 12, 15, 20, 30, 60, ∞, for which our table provides by means of the arguments in brackets. Using the Lagrangian multipliers, and the tabular entries corresponding to the bracketed column headings in row (54), we have the values alongside.

$$\begin{aligned} z(12, 54) = &+0{\cdot}7744 \times 0{\cdot}003\,23 \\ &-0{\cdot}7122 \times 0{\cdot}024\,40 \\ &+0{\cdot}6496 \times 0{\cdot}082\,10 \\ &-0{\cdot}5864 \times 0{\cdot}167\,43 \\ &+0{\cdot}5224 \times 0{\cdot}266\,84 \\ &+0{\cdot}4574 \times 0{\cdot}853\,88 \\ &-0{\cdot}3908 \times 0{\cdot}014\,23 \\ = &\ 0{\cdot}4647 \end{aligned}$$

TABLE OF PERCENTAGE POINTS OF THE χ^2 DISTRIBUTION

CALCULATED BY CATHERINE M. THOMPSON

EDITORIAL

In the calculation of the percentage points of the incomplete B-function $I_x(p, q)$ for large values of p and q described in the preceding paper, use has been made of the relation between this function and the incomplete Γ-function $I(u, p^*)$† which was tabulated by Karl Pearson (2). Since the incomplete Γ-function is related to the probability integral of χ^2, it was decided that the calculation of the percentage points of u already carried out should be extended to form complete tables of percentage points of χ^2. Before describing the method of computation used in deriving these tables it is desirable to relate them to existing tables and to define the relation between the functions.

In common terminology the probability distribution of χ^2 having ν degrees of freedom may be written

$$f(\chi^2) = \frac{(\frac{1}{2})^{\frac{1}{2}\nu}}{\Gamma(\frac{1}{2}\nu)} (\chi^2)^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}\chi^2}. \tag{1}$$

The probability integral of χ^2 or the chance that this quantity exceeds a given value χ^2 is then

$$P = P_\nu(\chi^2) = \int_{\chi^2}^{\infty} f(\chi^2)\, d\chi^2. \tag{2}$$

Conversely, for given degrees of freedom ν, the integral (2) will be equal to a given probability level P for one particular lower limit χ^2. This lower limit is called the percentage point χ^2 corresponding to ν and P; it will be denoted by $\chi^2_\nu(P)$.

These percentage points $\chi^2_\nu(P)$ were first tabulated by R. A. Fisher (1) for $P = 0{\cdot}99, 0{\cdot}98, 0{\cdot}95, 0{\cdot}90, 0{\cdot}80, 0{\cdot}70, 0{\cdot}50, 0{\cdot}30, 0{\cdot}20, 0{\cdot}10, 0{\cdot}05, 0{\cdot}02, 0{\cdot}01$‡ and for $\nu = 1\,(1)\,30$. Most entries in the body of Fisher's table are given to three decimal accuracy (i.e. four to five significant figures) but more decimals are given for small percentage points which are given to three-figure accuracy. In the table which follows $\chi^2_\nu(P)$ is tabulated for $P = 0{\cdot}995, 0{\cdot}99, 0{\cdot}975, 0{\cdot}95, 0{\cdot}90, 0{\cdot}75, 0{\cdot}50, 0{\cdot}25, 0{\cdot}10, 0{\cdot}05, 0{\cdot}025, 0{\cdot}01, 0{\cdot}005$ (which are the levels used for $I_x(p, q)$) whilst the range of the degrees of freedom has been extended up to $\nu = 100$. The percentage points are given to six significant figures, although for $\nu > 50$ the sixth figure may be in error by one or two units.

† Karl Pearson's notation was $I(u, p)$. To avoid confusion with the parameter p in $I_x(p, q)$ we have added the asterisk.

‡ In a later edition the level $P = 0{\cdot}001$ was added.

Table of percentage points of the χ^2 distribution

ν \ P	0·995	0·990	0·975	0·950	0·900	0·750
1	392704.10^{-10}	157088.10^{-9}	982069.10^{-9}	393214.10^{-8}	0·0157908	0·1015308
2	0·0100251	0·0201007	0·0506356	0·102587	0·210720	0·575364
3	0·0717212	0·114832	0·215795	0·351846	0·584375	1·212534
4	0·206990	0·297110	0·484419	0·710721	1·063623	1·92255
5	0·411740	0·554300	0·831211	1·145476	1·61031	2·67460
6	0·675727	0·872085	1·237347	1·63539	2·20413	3·45460
7	0·989265	1·239043	1·68987	2·16735	2·83311	4·25485
8	1·344419	1·646482	2·17973	2·73264	3·48954	5·07064
9	1·734926	2·087912	2·70039	3·32511	4·16816	5·89883
10	2·15585	2·55821	3·24697	3·94030	4·86518	6·73720
11	2·60321	3·05347	3·81575	4·57481	5·57779	7·58412
12	3·07382	3·57056	4·40379	5·22603	6·30380	8·43842
13	3·56503	4·10691	5·00874	5·89186	7·04150	9·29906
14	4·07468	4·66043	5·62872	6·57063	7·78953	10·1653
15	4·60094	5·22935	6·26214	7·26094	8·54675	11·0365
16	5·14224	5·81221	6·90766	7·96164	9·31223	11·9122
17	5·69724	6·40776	7·56418	8·67176	10·0852	12·7919
18	6·26481	7·01491	8·23075	9·39046	10·8649	13·6753
19	6·84398	7·63273	8·90655	10·1170	11·6509	14·5620
20	7·43386	8·26040	9·59083	10·8508	12·4426	15·4518
21	8·03366	8·89720	10·28293	11·5913	13·2396	16·3444
22	8·64272	9·54249	10·9823	12·3380	14·0415	17·2396
23	9·26042	10·19567	11·6885	13·0905	14·8479	18·1373
24	9·88623	10·8564	12·4011	13·8484	15·6587	19·0372
25	10·5197	11·5240	13·1197	14·6114	16·4734	19·9393
26	11·1603	12·1981	13·8439	15·3791	17·2919	20·8434
27	11·8076	12·8786	14·5733	16·1513	18·1138	21·7494
28	12·4613	13·5648	15·3079	16·9279	18·9392	22·6572
29	13·1211	14·2565	16·0471	17·7083	19·7677	23·5666
30	13·7867	14·9535	16·7908	18·4926	20·5992	24·4776
40	20·7065	22·1643	24·4331	26·5093	29·0505	33·6603
50	27·9907	29·7067	32·3574	34·7642	37·6886	42·9421
60	35·5346	37·4848	40·4817	43·1879	46·4589	52·2938
70	43·2752	45·4418	48·7576	51·7393	55·3290	61·6983
80	51·1720	53·5400	57·1532	60·3915	64·2778	71·1445
90	59·1963	61·7541	65·6466	69·1260	73·2912	80·6247
100	67·3276	70·0648	74·2219	77·9295	82·3581	90·1332
y_P	−2·5758	−2·3263	−1·9600	−1·6449	−1·2816	−0·6745

For $30 < \nu < 100$ interpolation formulae (6) or (7) of the Introduction may be used.

TABLE OF PERCENTAGE POINTS OF THE χ^2 DISTRIBUTION (*continued*)

ν \ P	0·500	0·250	0·100	0·050	0·025	0·010	0·005
1	0·454937	1·32330	2·70554	3·84146	5·02389	6·63490	7·87944
2	1·38629	2·77259	4·60517	5·99147	7·37776	9·21034	10·5966
3	2·36597	4·10835	6·25139	7·81473	9·34840	11·3449	12·8381
4	3·35670	5·38527	7·77944	9·48773	11·1433	13·2767	14·8602
5	4·35146	6·62568	9·23635	11·0705	12·8325	15·0863	16·7496
6	5·34812	7·84080	10·6446	12·5916	14·4494	16·8119	18·5476
7	6·34581	9·03715	12·0170	14·0671	16·0128	18·4753	20·2777
8	7·34412	10·2188	13·3616	15·5073	17·5346	20·0902	21·9550
9	8·34283	11·3887	14·6837	16·9190	19·0228	21·6660	23·5893
10	9·34182	12·5489	15·9871	18·3070	20·4831	23·2093	25·1882
11	10·3410	13·7007	17·2750	19·6751	21·9200	24·7250	26·7569
12	11·3403	14·8454	18·5494	21·0261	23·3367	26·2170	28·2995
13	12·3398	15·9839	19·8119	22·3621	24·7356	27·6883	29·8194
14	13·3393	17·1170	21·0642	23·6848	26·1190	29·1413	31·3193
15	14·3389	18·2451	22·3072	24·9958	27·4884	30·5779	32·8013
16	15·3385	19·3688	23·5418	26·2962	28·8454	31·9999	34·2672
17	16·3381	20·4887	24·7690	27·5871	30·1910	33·4087	35·7185
18	17·3379	21·6049	25·9894	28·8693	31·5264	34·8053	37·1564
19	18·3376	22·7178	27·2036	30·1435	32·8523	36·1908	38·5822
20	19·3374	23·8277	28·4120	31·4104	34·1696	37·5662	39·9968
21	20·3372	24·9348	29·6151	32·6705	35·4789	38·9321	41·4010
22	21·3370	26·0393	30·8133	33·9244	36·7807	40·2894	42·7956
23	22·3369	27·1413	32·0069	35·1725	38·0757	41·6384	44·1813
24	23·3367	28·2412	33·1963	36·4151	39·3641	42·9798	45·5585
25	24·3366	29·3389	34·3816	37·6525	40·6465	44·3141	46·9278
26	25·3364	30·4345	35·5631	38·8852	41·9232	45·6417	48·2899
27	26·3363	31·5284	36·7412	40·1133	43·1944	46·9630	49·6449
28	27·3363	32·6205	37·9159	41·3372	44·4607	48·2782	50·9933
29	28·3362	33·7109	39·0875	42·5569	45·7222	49·5879	52·3356
30	29·3360	34·7998	40·2560	43·7729	46·9792	50·8922	53·6720
40	39·3354	45·6160	51·8050	55·7585	59·3417	63·6907	66·7659
50	49·3349	56·3336	63·1671	67·5048	71·4202	76·1539	79·4900
60	59·3347	66·9814	74·3970	79·0819	83·2976	88·3794	91·9517
70	69·3344	77·5766	85·5271	90·5312	95·0231	100·425	104·215
80	79·3343	88·1303	96·5782	101·879	106·629	112·329	116·321
90	89·3342	98·6499	107·565	113·145	118·136	124·116	128·299
100	99·3341	109·141	118·498	124·342	129·561	135·807	140·169
y_P	0·0000	+0·6745	+1·2816	+1·6449	+1·9600	+2·3263	+2·5758

For $\nu > 100$ take $\chi^2_\nu(P) = \nu\left\{1-\frac{2}{9\nu}+y_P\sqrt{\frac{2}{9\nu}}\right\}^3$ or $\chi^2_\nu(P)=\frac{1}{2}\{y_P+\sqrt{(2\nu-1)}\}^2$, according to the degree of accuracy required.

The relation with the incomplete Γ-function is given by

$$P_\nu(\chi^2) = 1 - I(u, p^*) \tag{3}$$

where
$$\chi^2 = 2u\sqrt{(p^*+1)}, \tag{4}$$

the degrees of freedom, ν, of χ^2 being given by

$$\nu = 2p^* + 2. \tag{5}$$

The above relations were used for the computation of $\chi^2_\nu(P)$. Most of the entries were obtained from the *Tables of the Incomplete Γ-Function* (2). In these tables the column headed $p^* = \frac{1}{2}\nu - 1$ was entered, the root u of

$$I(u, p^*) = 1 - P$$

found by inverse interpolation and transformed into the corresponding percentage point for χ^2 by substitution in equation (4).

Although for small values of p^* and u the table of $I(u, p^*)$ is not interpolable, formal inverse interpolation in this range of the table still yields approximate values of the percentage points. To make these accurate, auxiliary tables of $P_\nu(\chi^2)$ were constructed for arguments χ^2 in the neighbourhood of the approximate percentage points, and the exact values of $\chi^2(P)$ found by inverse interpolation in these auxiliary tables. The latter were constructed from the expansion of $P_\nu(\chi^2)$ given on p. xxxi of *Tables for Statisticians and Biometricians*, Part I (3). Since the auxiliary tables were required only for small values of χ^2 a few terms in the expansions were sufficient to yield $P_\nu(\chi^2)$ to the required accuracy.

Whilst the existing table of percentage points of χ^2 (1) is confined to the range $\nu = 1\,(1)\,30$, the range $\nu = 30\,(10)\,100$ has been added in the table below. This has been done because the customary approximation to $\chi^2_\nu(P)$ by the corresponding normal deviates is not very satisfactory in this range of ν. There are, in fact, percentage points which differ from the approximate ones in the second significant figure. In our table, however, linear interpolation (which is particularly convenient at interval 10) yields interpolates accurate to about four significant figures. If we write $\nu = 10k + m$ with $3 \leqslant k < 10$ and $0 \leqslant m < 10$ then we have

$$\chi^2_\nu = \tfrac{1}{10}\{(10-m)\,\chi^2_{10k} + m\chi^2_{10k+10}\}. \tag{6}$$

For instance for $\nu = 54$ and $P = 0{\cdot}01$ we have

$$\chi^2_{54}(0{\cdot}01) = \tfrac{1}{10}\{6\chi^2_{50}(0{\cdot}01) + 4\chi^2_{60}(0{\cdot}01)\} = 81{\cdot}04.$$

If higher accuracy is required we have to use the four point Lagrangian formula, viz.

$$\chi^2_\nu = L_{-1}\chi^2_{10k-10} + L_0\chi^2_{10k} + L_1\chi^2_{10k+10} + L_2\chi^2_{10k+20} \tag{7}$$

where the Lagrangian coefficients L_{-1}, L_0, L_1 and L_2 are tabulated below.

m	L_{-1}	L_0	L_1	L_2	
	$-$	$+$	$+$	$-$	
0	0·0000	1·0000	0·0000	0·0000	10
1	0·0285	0·9405	0·1045	0·0165	9
2	0·0480	0·8640	0·2160	0·0320	8
3	0·0595	0·7735	0·3315	0·0455	7
4	0·0640	0·6720	0·4480	0·0560	6
5	0·0625	0·5625	0·5625	0·0625	5
	$-$	$+$	$+$	$-$	
	L_2	L_1	L_0	L_{-1}	m

Returning to the above example we obtain

$$\begin{aligned}\chi^2_{54}(0{\cdot}01) &= -0{\cdot}0640\chi^2_{40}(0{\cdot}01)+0{\cdot}6720\chi^2_{50}(0{\cdot}01)+0{\cdot}4480\chi^2_{60}(0{\cdot}01)-0{\cdot}0560\chi^2_{70}(0{\cdot}01)\\ &= 81{\cdot}069\end{aligned}$$

which is the exact interpolate to five figures.

For $\nu > 100$ we can make use of Fisher's approximation to $P_\nu(\chi^2)$ by the normal probability integral and calculate $\chi^2_\nu(P)$ from the formula

$$\chi^2_\nu(P) = \tfrac{1}{2}\{y_P+\sqrt{(2\nu-1)}\}^2 \tag{8}$$

where the normal deviates y_P corresponding to the thirteen percentage levels are given in the last line of the main table. For $\nu = 100$, this approximation has an accuracy of about 1 % in the worst cases, but as ν increases it becomes more accurate.

A more accurate approximation has been given by Wilson and Hilferty(4); this assumes that $(\chi^2/\nu)^{\frac{1}{3}}$ is normally distributed about $1-2/(9\nu)$ with standard deviation $\sqrt{(2/9\nu)}$. That is to say, the probability levels may be calculated from

$$\chi^2_\nu(P) = \nu\left\{1-\frac{2}{9\nu}+y_P\sqrt{\frac{2}{9\nu}}\right\}^3. \tag{9}$$

Comparative numerical values showing the relative accuracy of these two formulæ are given in the Note by Mrs M. Merrington on pp. 200–2 below.

REFERENCES

(1) FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

(2) PEARSON, K. (1922). *Tables of the Incomplete Γ-function.* London: *Biometrika.*

(3) PEARSON, K. (1930). *Tables for Statisticians and Biometricians.* Part I. London: *Biometrika.*

(4) WILSON, E. B. and HILFERTY, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, **17**, 684.

MISCELLANEA

(i) **Theory of Probability.** By HAROLD JEFFREYS. Oxford University Press. 1939. 7+380 pp. 21*s*. net.

In the history of the application of probability theory to the problem of drawing inferences from observations, no set of ideas has played a more controversial role than that associated with inverse probability. Various leaders of modern thought in statistical inference have pointed out the logical difficulties inherent in the application of inverse probability. R. A. Fisher, through his methods of maximum likelihood and of fiducial limits, has introduced principles of statistical inference which make the introduction of the notion of inverse probability irrelevant. These principles have been extended and refined by J. Neyman, E. S. Pearson, A. Wald and others, until now we have available in statistical literature a self-consistent discipline of statistical inference which is independent of inverse probability.

In the present book the author proposes a system of statistical inference based on the principles of inverse probability, applying it to the same problems which have been treated by Fisher, Neyman, Pearson and others without using inverse probability. The attitude which the author takes towards probability theory is somewhat similar to that taken by J.M. Keynes. Probability is regarded as a subjective phenomenon. The essential idea is that probability is a matter of comparing 'reasonable degrees of belief' in propositions. In Chapter I the author goes through a considerable amount of psychological and philosophical discussion attempting to justify this approach. This discussion is finally formalized by a set of six axioms. The primitive or undefined notion is that of the relation 'given p, q is more probable than r', where p, q, and r are propositions. The symbol used for denoting the probability of q given p is $P(q \mid p)$. The six axioms which are used are as follows:

(1) Given p, q is either more or less probable than r, or both are equally probable; and no two of these alternatives can be true.

(2) If p, q, r, s are four propositions and given p, q is more probable than r and r is more probable than s, then given p, q is more probable than s.

(3) All propositions deducible from a proposition p have the same probability on data p; and all propositions inconsistent with p have the same probability on data p.

(4) If, given p, q and q' cannot both be true, and if, given p, r and r' cannot both be true, and if, given p, q and r are equally probable and q' and r' are equally probable, then given p, 'q or q'' and 'r or r'' are equally probable.

(5) The set of possible probabilities on given data, ordered in terms of the relation 'more probable', is not of higher ordinal type than the continuum including the end-points.

(6) If pq entails r, then $P(qr \mid p) = P(q \mid p)$.

In axiom 6, the expression 'a entails b' is defined as meaning 'a is deducible from b', or 'a is identical with b', or 'a is identical with some proposition asserted in b'. The expression 'ab' is taken as the logical product, that is 'both a and b'.

The introduction of numbers for expressing probabilities is made through three 'conventions'. The first convention associates the larger of two numbers with the more probable of two propositions, the second one states that if given p, q and q' are mutually exclusive, then $P(q \mid p) + P(q' \mid p) = P(q \text{ or } q' \mid p)$, and the third states that if p entails q, then $P(q \mid p) = 1$.

In order to be thoroughly rigorous in his axiomatic approach, presumably the author should have postulated the existence of an aggregate of propositions on which to operate. The possibility of an infinite number of propositions should, of course, not be excluded, as will be seen when the author applies his theory to problems involving continuous random variables. In the case of an infinite number of propositions, it appears that the assumption should be made that axiom 4 would hold in case of two infinite sets of mutually exclusive alternatives. A similar assumption would have to be made for the second convention. The proponent of the measure theory approach to probability has essentially the same problem to deal with, but he handles it by assuming the existence of set functions which are completely additive over his postulated field of sets.

Proceeding from his six axioms and three conventions the author completes his chapter by establishing twelve theorems which are used as the basis for the work in the subsequent chapters. The question of an infinite number of alternatives is not covered by these theorems.

Chapter II, on 'Direct Probabilities', is devoted to derivations and discussions of the binomial, normal, Poisson, Pearson, multinomial, Chi-square, t, z and other frequency laws and their properties. The characteristic function and illustrations of its use in the determination of probability laws are presented.

In building up his system of statistical inference Jeffreys proceeds in Chapters III–VIII by applying the principle of inverse probability to the results of Chapter II. Thus if S is a set of observations subject to a given discrete distribution law $P(S \mid \theta H)$ derived under a distribution hypothesis H, where θ is a parameter to be estimated, an *a priori* probability function $g(\theta)\,d\theta$ for θ is introduced. The *posterior* probability law of θ given S, say $P(\theta \mid SH)$, is given by

$$P(\theta \mid SH)\,d\theta = \frac{P(S \mid \theta H)\,g(\theta)\,d\theta}{\int \Sigma P(S \mid \theta H)\,g(\theta)\,d\theta},$$

where Σ denotes summation with respect to all possible configurations of values of S, and the integral is taken with respect to θ. A similar analysis results when S is subject to a continuous distribution law. The function $P(\theta \mid SH)$ is then taken as a basis for estimating θ. For two values of θ, say θ_1 and θ_2, θ_1 is called more probable than θ_2 if $P(\theta_1 \mid SH) > P(\theta_2 \mid SH)$. Needless to say, this comparison of *posterior* probabilities depends on the choice of $g(\theta)$. Now, from the point of view of applying this discipline, $g(\theta)$ would rarely, if ever, be known, and the controversy over inverse probability centres around the problem of choosing $g(\theta)$. The author adopts two rules for selecting $g(\theta)$: (1) If the parameter may have any value in a finite range, or from $-\infty$ to $+\infty$ its *prior* probability should be taken as uniformly distributed. (2) If the parameter may conceivably have any value from 0 to $+\infty$, the *prior* probability of its logarithm should be taken as uniformly distributed. The adoption of these two rules appears to the reviewer to be extremely vulnerable. First of all, what does it mean, in general, for a parameter to be uniformly distributed on the interval $-\infty$ to $+\infty$? This question appears to be particularly in order since the author is performing the formal calculus of probabilities in exactly the same manner in which measure proponents calculate probabilities. They continually use the property that a finite total probability (taken arbitrarily as unity) is associated with each probability function. Presumably, meaning could be injected by carrying out the work for a finite interval $-K$ to K and then taking the limit of the results or answers as $K \to \infty$. Similarly, it may be asked what it means to have the logarithm of a parameter uniformly distributed on a semi-infinite interval. Owing to the nature of the particular problems to which Jeffreys applies these rules, it happens that $P(S \mid \theta H)$ is such that formal difficulties of convergence do not arise in obtaining $P(\theta \mid SH)$. For the finite interval, why should one choose the parameter to be uniformly distributed rather than the square or some other function of the parameter, or for the semi-infinite interval $(0, \infty)$ why should one choose the logarithm of the parameter rather than some other function to be uniformly distributed? It is easy to show that the assumption of uniform distribution is, in general, inconsistent with that of uniform distribution of any single-valued function of the parameter.

Chapter IV, entitled 'Approximate Methods and Simplifications', contains discussions, some of them rather heuristic, of the problem of estimation involved in such topics as maximum likelihood, least squares, errors due to grouping, rank correlation, contingency, artificial randomization, etc. The inverse probability approach is, of course, maintained throughout.

The problem of significance tests is treated in Chapters V and VI. Jeffreys' attitude toward significance tests follows along the lines of his concept of probability and consists in comparing *posterior* probabilities. More specifically, suppose θ is a parameter and it is desired to test the hypothesis that $\theta = 0$ on the basis of a given set of observations and an hypothesis H regarding the distribution law of S for given θ. Let q denote the hypothesis that $\theta = 0$, and

$\bar{q}$ denote the hypothesis that θ has some other value. Jeffreys' criterion for making the significance test is the ratio K of the two *posterior* probabilities $P(q \mid SH)$ and $P(\bar{q} \mid SH)$. The value of K itself, and not the probability integral of K under the *null* hypothesis q, is proposed as the criterion. Expressions for K are found for such problems as contingency, comparison of means and variances in samples, consistency of two Poisson parameters, correlation—problems which have already been treated by Fisher, Neyman, Pearson and others by other approaches free from inverse probability. In the treatment of all of these problems, the author arrives at four principal forms of K, which he proceeds to tabulate in an appendix for various values of sample size and for five grades of significance corresponding to $K = 1, 10^{-\frac{1}{2}}, 10^{-1}, 10^{-\frac{3}{2}}, 10^{-2}$. The last two chapters, i.e. VII and VIII, are entitled 'Frequency Definitions and Direct Methods', and 'General Questions', respectively. These chapters are primarily philosophical excursions undertaken in an attempt to show that no existing definition of probability avoids the notion of 'degrees of reasonable belief', and to justify his own approach as well as his attitude toward inverse probability. The discussion is almost entirely informal and non-mathematical and as such it must be regarded in the category of personal opinion.

The book lacks strict mathematical rigour in various places, but from the point of view of general flow of discussion it is interestingly written. It contains many keenly chosen quotations and side remarks charged with a delightfully subtle humour which has characterized the author in other books.

From a scientific point of view it is doubtful that there will be many scholars thoroughly familiar with the system of statistical inference initiated by R. A. Fisher and extended by J. Neyman, E. S. Pearson, A. Wald and others who will abandon this system in favour of the one proposed by Jeffreys in which inverse probability plays the central role.

PRINCETON UNIVERSITY — S. S. WILKS

(ii) **A Bibliography of Human Morphology, 1914–1939.** By WILTON M. KROGMAN. United States of America: University of Chicago Press; Great Britain and Ireland: Cambridge University Press. 1941. Price 18*s*.

The title of this volume may mislead to some extent. The 11,000 odd references in it were collected to aid physical anthropologists and the work will be of greater value to them than to other research workers, such as anatomists and geneticists, who are concerned with human morphology. The non-German literature is said to be covered more thoroughly than in the second edition of Rudolf Martin's *Lehrbuch*, and there is no other comprehensive bibliography of the subject for the period since 1928. It is not claimed that the list is exhaustive, and the most stringent selection appears to have been made in the section on blood groups for which fuller bibliographies are available.

G. M. M.

(iii) A Property of the Distribution of Extremes

BY H. E. DANIELS

Wool Industries Research Association

If the chance of an observation being less than x is P, then P^n is the chance that the greatest of a random sample of n is less than x. The constants of the distribution of the greatest of a sample in the important case when $P = \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} \frac{dx}{\sqrt{(2\pi)}}$ have been calculated by Tippett (1925) for values of n up to 1000, and in a paper in which all possible limiting forms of P^n are discussed, Fisher & Tippett (1928) give limiting formulae from which approximate values of the constants are calculated for large samples.

In the present note attention is drawn to a curious approximate relation connecting the mean M and standard deviation σ of this distribution which holds with high accuracy for all values of n. It was arrived at empirically and appears to have no obvious mathematical derivation. The formula is

$$M = 2\cot \tfrac{1}{2}\pi\sigma.$$

The values of $2\cot\frac{1}{2}\pi\sigma$, calculated from Tippett's figures, are compared in Table 1 with Tippett's values of M. The greatest discrepancy, $\delta = M - 2\cot\frac{1}{2}\pi\sigma$, over the range of n up to 1000 occurs at $n = 10$ where it is no more than about $1\frac{1}{2}$ %. For n greater than 1000 the penultimate limiting values given in Table A of Fisher & Tippett's paper are used in Table 2 and the discrepancies are again found to be small. The most serious is of the order of 3 % when n is 7228, but their Table B suggests that the penultimate limiting values of M and σ are probably underestimated and as $2\cot\frac{1}{2}\pi\sigma$ is fairly sensitive to changes of σ in the region of $\sigma = 0{\cdot}3$, the real discrepancy may perhaps be smaller.

Table 1

n	M	σ	$2\cot\frac{1}{2}\pi\sigma$	$\delta = M - 2\cot\frac{1}{2}\pi\sigma$
1	0·0000	1·0000	0·0000	0·0000
2	0·5642	0·8257	0·5617	−0·0025
5	1·1630	0·6690	1·1448	−0·0182
10	1·5388	0·5868	1·5176	−0·0212
20	1·8675	0·5251	1·8483	−0·0192
60	2·3193	0·4545	2·3086	−0·0107
100	2·5076	0·4294	2·5012	−0·0064
200	2·7460	0·4009	2·7448	−0·0012
500	3·0367	0·3704	3·0407	0·0040
1000	3·2414	0·3514	3·2475	0·0061

Table 2. *Penultimate approximate values*

n	M	σ	$2\cot\frac{1}{2}\pi\sigma$	$\delta = M - 2\cot\frac{1}{2}\pi\sigma$
7228	3·7697	0·3039	3·8661	0·0964
637×10^3	4·7719	0·2499	4·8311	0·0592
264×10^9	6·9262	0·1787	6·9369	0·0107

Fisher and Tippett show that in the ultimate limiting form of the distribution the mode m, mean M and standard deviation σ are related by the formulae

$$M = m + \gamma c, \quad \sigma^2 = \tfrac{1}{6}\pi^2 c^2,$$

where $c = m/(m^2+1)$ and $\gamma = 0{\cdot}577216$ is Euler's constant. Consequently

$$M = \frac{1}{2c} + \sqrt{\left(\frac{1}{4c^2} - 1\right)} + \gamma c \sim \frac{1}{c} = \frac{\pi}{\sigma\sqrt{6}} = \frac{1{\cdot}282}{\sigma}$$

as σ becomes small with increasing n. On the other hand, our approximate relation gives

$$M = 2\cot\tfrac{1}{2}\pi\sigma \sim \frac{4}{\pi\sigma} = \frac{1{\cdot}274}{\sigma}.$$

The error at $n = \infty$ is thus seen to be less than 1 %.

REFERENCES

Tippett, L. H. C. (1925). *Biometrika*, **17**, 364.

Fisher, R. A. & Tippett, L. H. C. (1928). *Proc. Camb. Phil. Soc.* **24**, 180.

(iv) Proof of Relations connected with the Tetrachoric Series and its Generalization

By M. G. KENDALL

If a bivariate normal distribution F with variates x_1 and x_2 and correlation ρ is doubly dichotomized at $x_1 = h$, $x_2 = k$, and

$$d = \int_h^\infty \int_k^\infty dF,$$

it is known that

$$d = \sum_{r=0}^{\infty} \rho^r \tau_r(h)\, \tau_r(k), \tag{1}$$

where τ_r is the rth tetrachoric function defined by

$$\tau_r(x) = \frac{H_{r-1}(x) f(x)}{(r!)^{\frac{1}{2}}}, \tag{2}$$

and $f(x)$ being the function $\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$ and $H_r(x)$ the rth Hermite polynomial defined by

$$H_r f(x) = \left(-\frac{d}{dx}\right)^r f(x) = (-D)^r f(x). \tag{3}$$

The purpose of this note is to present a simple proof of this result,* to prove that the series of equation (1) is convergent for $|\rho| \leqslant 1$ and to generalize the series to the case of the multivariate normal distribution.†

The characteristic function of

$$dF = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right\}$$

is, by definition,

$$\phi(t_1, t_2) = \int_{-\infty}^\infty \int_{-\infty}^\infty \exp\{it_1 x_1 + it_2 x_2\}\, dF,$$

and is easily seen by direct integration to be equal to

$$\exp\{-\tfrac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2)\}. \tag{4}$$

We have then, for all finite t_1, t_2,

$$\phi(t_1, t_2) = \exp(-\tfrac{1}{2}t_1^2) \exp(-\tfrac{1}{2}t_2^2) \sum_{r=0}^{\infty} (-\rho)^r \frac{t_1^r t_2^r}{r!}.$$

Now

$$d = \int_h^\infty \int_k^\infty dF = \frac{1}{(2\pi)^2} \int_h^\infty dx_1 \int_k^\infty dx_2 \int_{-\infty}^\infty \int_{-\infty}^\infty \phi(t_1, t_2) \exp\{-it_1 x_1 - it_2 x_2\}\, dt_1\, dt_2.$$

Substituting for $\phi(t_1, t_2)$ we have, for the coefficient of $(-\rho)^r/r!$ in this expression,

$$\frac{1}{4\pi^2} \int_h^\infty dx_1 \int_k^\infty dx_2 \int_{-\infty}^\infty \int_{-\infty}^\infty \exp(-\tfrac{1}{2}t_1^2) \exp(-\tfrac{1}{2}t_2^2)\, t_1^r t_2^r \exp\{-it_1 x_1 - it_2 x_2\}\, dt_1\, dt_2, \tag{5}$$

and this is the product of two integrals, the first of which is

$$\frac{1}{2\pi} \int_h^\infty dx_1 \int_{-\infty}^\infty \exp(-\tfrac{1}{2}t_1^2)\, t_1^r \exp\{-it_1 x_1\}\, dt_1 \tag{6}$$

and the second of which is a similar expression in x_2 and t_2.

* The expansion (1) appears to have been given for the first time by G. Mehler, 'Reihewentwicklung nach Laplaceschen Functionen höherer Ordnung', *J. reine angew. Math.* **66**, 161.

† See Note at end of paper.

Now, since
$$\int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}t^2)\, t^r e^{-itx}\, dt = \frac{\partial^r}{\partial(-ix)^r}\int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}t^2)\, e^{-itx}\, dt$$
$$= \sqrt{(2\pi)}\, i^r D^r e^{-\frac{1}{2}x^2}$$
$$= 2\pi(-i)^r H_r(x) f(x), \tag{7}$$

(6) is equal to
$$(-i)^r \int_h^{\infty} dx\, H_r(x) f(x) = (-i)^r H_{r-1}(h) f(h),$$

and thus (5) is equal to
$$(-1)^r H_{r-1}(h) f(h) H_{r-1}(k) f(k),$$

and hence
$$d = \Sigma \frac{\rho^r}{r!} H_{r-1}(h) f(h) H_{r-1}(k) f(k)$$
$$= \Sigma \rho^r \tau_r(h)\, \tau_r(k), \tag{8}$$
the tetrachoric series.

Now for the convergence of the series, consider
$$|\tau_r(h)\, \tau_r(k)| = (r!)^{-1} |H_{r-1}(h) f(h) H_{r-1}(k) f(k).$$

From (6)
$$|H_{r-1}(h) f(h)| \leqslant \frac{1}{2\pi}\left|\int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}t^2)\, t^{r-1} e^{-ith}\, dt\right|$$
$$\leqslant \frac{1}{\pi}\int_0^{\infty} \exp(-\tfrac{1}{2}t^2)\, t^{r-1}\, dt$$
$$\leqslant \frac{2^{\frac{r-2}{2}}}{\pi}\, \Gamma\left(\frac{r}{2}\right).$$

Hence
$$|\tau_r(h)\, \tau_r(k)| \leqslant \frac{2^{r-2}\Gamma^2\left(\dfrac{r}{2}\right)}{\pi r!}$$
$$\leqslant \sim \frac{2^{r-2}\, e^{-(r-2)}\, 2\pi\left(\dfrac{r-2}{2}\right)^{r-1}}{\pi\sqrt{(2\pi)}\, e^{-r} r^{r+\frac{1}{2}}}$$
$$\leqslant \sim \frac{1}{\sqrt{(2\pi)}\, r^{\frac{3}{2}}}.$$

Thus the tetrachoric series converges for $|\rho| \leqslant 1$, though possibly slowly near $|\rho| = 1$.

Now consider the general multivariate normal distribution
$$dF = \frac{1}{(2\pi)^{\frac{n}{2}} R^{\frac{1}{2}}} \exp -\frac{1}{2R}\{\Sigma R_{jj} x_j^2 + 2\Sigma R_{jk} x_j x_k\}\, dx_1 \ldots dx_n,$$

where
$$R = \begin{vmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{12} & 1 & \ldots & \rho_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{1n} & \rho_{2n} & \ldots & 1 \end{vmatrix}$$

and R_{jk} is the minor of the jth row and kth column.

The characteristic function is
$$\phi(t_1, \ldots, t_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} dF \exp(\Sigma i t_j x_j)\, dx_1 \ldots dx_n.$$

To evaluate this integral make the transformation

$$x_j = \sum_k \alpha_{jk} \xi_k$$

and choose the α's so as to reduce the second degree terms in ξ in the exponent to the canonical form $\Sigma\xi^2$. The remaining terms in t and ξ will obviously be linear in each. We may then make a further transformation $\xi' = \xi -$(linear function of the t's) and the remaining terms, apart from $\Sigma\xi'^2$, will be a quadratic in the t's, and equal, say, to

$$\Sigma\gamma_j t_j^2 + 2\Sigma\delta_{jk} t_j t_k.$$

The integration abolishes the terms in ξ' and we find that the characteristic function is proportional to

$$\exp\{\Sigma\gamma_j t_j^2 + 2\Sigma\delta_{jk} t_j t_k\}.$$

Putting all but two of the t's zero we see by comparison with (4) that the terms in t_j, t_k is $-\frac{1}{2}(t_j^2 + t_k^2 + 2\rho_{jk} t_j t_k)$ and thus the characteristic function is

$$\exp\{-\tfrac{1}{2}(\Sigma t_j^2 + 2\Sigma\rho_{jk} t_j t_k)\}. \tag{9}$$

The generalized tetrachoric expansion can be obtained by an expansion of (9) in terms of the ρ's and the application of the foregoing procedure. For instance, with three variates we have

$$\begin{aligned}\exp\Sigma(-\rho_{jk} t_j t_k) &= \Sigma\frac{(-1)^r}{r!}(\rho_{12}t_1t_2 + \rho_{23}t_2t_3 + \rho_{13}t_1t_3)^r \\ &= \Sigma(-1)^r \frac{\rho_{12}^j \rho_{23}^k \rho_{13}^l}{j!\,k!\,l!} t_1^{j+k} t_2^{j+l} t_3^{k+l},\end{aligned}$$

and on integration

$$\begin{aligned}d &= \int_{h_1}^{\infty}\int_{h_2}^{\infty}\int_{h_3}^{\infty} dF \\ &= \Sigma \frac{\rho_{12}^j \rho_{23}^k \rho_{13}^l}{j!\,k!\,l!} H_{j+k-1}(h_1) f(h_1) H_{j+l-1}(h_2) f(h_2) H_{k+l-1}(h_3) f(h_3),\end{aligned}$$

which will also be found to be convergent.

[*Note.* Since writing this paper, under the impression that the results were new, I am indebted to Dr A. C. Aitken for pointing out that similar results have been given by him in lectures for a number of years. Among published work, reference may be made to:

(*a*) P. 175 of Aitken & Turnbull's *Theory of Canonical Matrices* (Blackie, 1931), where a more direct method of deriving the characteristic function of the multivariate normal distribution has been given;

(*b*) A paper on 'Fourfold sampling with and without replacement' by A. C. Aitken & H.T. Gonin (1935, *Proc. Roy. Soc. Edinb.* **55**, 114), where the tetrachoric expansions are discussed and new series associated with the correlated binomial and the correlated hypergeometric distributions are derived.

As the work of the Edinburgh school on this subject may not, however, be generally familiar, the Editor has suggested that this short paper should be published, together with the foregoing references, in order to bring the recent developments before a wider statistical audience. M.G.K.]

(v) The Cumulants of the Distribution of the Square of a Variate

BY J. B. S. HALDANE, F.R.S.

The following problem has arisen in several biometric investigations. The cumulants of the distribution of x are known, and it is desired to find the cumulants of the distribution of x^2. As this problem is likely to arise in future, it seems desirable to give the appropriate transformations for the first few cumulants.

Let $\kappa_1, \kappa_2, \kappa_3, \ldots$ be the cumulants of x.

Let $\mu'_1, \mu'_2, \mu'_3, \ldots$ be the moments of x^2 about zero.

Let $\mu_2, \mu_3, \mu_4, \ldots$ be the moments of x^2 about its mean.

Let $\kappa'_1, \kappa'_2, \kappa'_3, \ldots$ be the cumulants of x^2.

Then μ'_r is the $2r$th moment of x. These have been given in terms of the cumulants up to the 10th, i.e. μ'_5, in the general case by Kendall (1940), and up to the 12th, i.e. μ'_6, by Haldane (1938) when $\kappa_1 = 0$. We consider the general case first. We have such expressions as

$$\mu'_2 = \kappa_1^4 + 6\kappa_1^2\kappa_2 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + \kappa_4.$$

From these we calculate the moments μ_r, and hence the cumulants. The results are:

$$\left.\begin{aligned}
\kappa'_1 &= \kappa_1^2 + \kappa_2,\\
\kappa'_2 &= 4\kappa_1^2\kappa_2 + 2(2\kappa_1\kappa_3 + \kappa_2^2) + \kappa_4,\\
\kappa'_3 &= 8\kappa_1^2(\kappa_1\kappa_3 + 3\kappa_2^2) + 4(3\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 2\kappa_2^3) + 2(3\kappa_1\kappa_5 + 6\kappa_2\kappa_4 + 5\kappa_3^2) + \kappa_6,\\
\kappa'_4 &= 16\kappa_1^2(\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 12\kappa_2^3)\\
&\quad + 16(2\kappa_1^3\kappa_5 + 18\kappa_1^2\kappa_2\kappa_4 + 12\kappa_1^2\kappa_3^2 + 36\kappa_1\kappa_2^2\kappa_3 + 3\kappa_2^4)\\
&\quad + 8(3\kappa_1^2\kappa_6 + 18\kappa_1\kappa_2\kappa_5 + 32\kappa_1\kappa_3\kappa_4 + 18\kappa_2^2\kappa_4 + 30\kappa_2\kappa_3^2)\\
&\quad + 8(\kappa_1\kappa_7 + 3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8.
\end{aligned}\right\} \quad (1)$$

After this the expressions become very heavy. When $\kappa_1 = 0$, i.e. x has its mean zero, most of the terms vanish, and we have

$$\left.\begin{aligned}
\kappa'_1 &= \kappa_2,\\
\kappa'_2 &= 2\kappa_2^2 + \kappa_4,\\
\kappa'_3 &= 8\kappa_2^3 + 2(6\kappa_2\kappa_4 + 5\kappa_3^2) + 6,\\
\kappa'_4 &= 48\kappa_2^4 + 48\kappa_2(3\kappa_2\kappa_4 + 5\kappa_3^2) + 8(3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8,\\
\kappa'_5 &= 384\kappa_2^5 + 960\kappa_2^2(\kappa_2\kappa_4 + 5\kappa_3^2) + 80(16\kappa_2\kappa_4^2 + 28\kappa_2\kappa_3\kappa_5 + 6\kappa_2^2\kappa_6 + 25\kappa_3^2\kappa_4)\\
&\quad + 2(20\kappa_2\kappa_8 + 60\kappa_3\kappa_7 + 100\kappa_4\kappa_6 + 63\kappa_5^2) + \kappa_{10},\\
\kappa'_6 &= 3840\kappa_2^6 + 9600\kappa_2^3(3\kappa_2\kappa_4 + 10\kappa_3^2) + 4800(2\kappa_2^2\kappa_6 + 14\kappa_2^2\kappa_3\kappa_5 + 8\kappa_2^2\kappa_4^2 + 25\kappa_2\kappa_3^2\kappa_4 + 3\kappa_3^4)\\
&\quad + 40(30\kappa_2^2\kappa_8 + 180\kappa_2\kappa_3\kappa_7 + 300\kappa_2\kappa_4\kappa_6 + 226\kappa_3^2\kappa_6 + 189\kappa_2\kappa_5^2 + 672\kappa_3\kappa_4\kappa_5 + 132\kappa_4^3)\\
&\quad + 4(15\kappa_2\kappa_{10} + 55\kappa_3\kappa_9 + 120\kappa_4\kappa_8 + 198\kappa_5\kappa_7 + 113\kappa_6^2) + \kappa_{12}.
\end{aligned}\right\} \quad (2)$$

Finally, if x be symmetrically distributed, so that all its odd cumulants vanish,

$$\left.\begin{aligned}
\kappa_1' &= \kappa_2,\\
\kappa_2' &= 2\kappa_2^2+\kappa_4,\\
\kappa_3' &= 8\kappa_2^3+12\kappa_2\kappa_4+\kappa_6,\\
\kappa_4' &= 48\kappa_2^4+144\kappa_2^2\kappa_4+8(3\kappa_2\kappa_6+4\kappa_4^2)+\kappa_8,\\
\kappa_5' &= 384\kappa_2^5+1920\kappa_2^3\kappa_4+160\kappa_2(3\kappa_2\kappa_6+8\kappa_4^2)+40(\kappa_2\kappa_8+5\kappa_4\kappa_6)+\kappa_{10},\\
\kappa_6' &= 3840\kappa_2^6+28800\kappa_2^4\kappa_4+9600\kappa_2^2(\kappa_2\kappa_6+4\kappa_4^2)+240(5\kappa_2^2\kappa_8+50\kappa_2\kappa_4\kappa_6+22\kappa_4^3)\\
&\quad +4(15\kappa_2\kappa_{10}+120\kappa_4\kappa_8+113\kappa_6^2)+\kappa_{12}.
\end{aligned}\right\} \tag{3}$$

I have bracketed together terms which are products of the same number of κ_r's. If x is a linear function of observed numbers in a sample of n, every κ_n is proportional to x, so the terms in brackets will all be multiples of the same power of n.

REFERENCES

HALDANE, J. B. S. (1938). 'The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small.' *Biometrika*, **29**, 389–91.

KENDALL, M. G. (1940). 'The derivation of multivariate sampling formulae from univariate formulae by symbolic operation.' *Ann. Eugen., Lond.*, **10**, 392–402.

(vi) Numerical approximations to the percentage points of the χ^2 distribution

BY MAXINE MERRINGTON

The use of two approximate formulae has been suggested for calculating a percentage point, $\chi^2_\nu(P)$, of the χ^2 distribution, corresponding to ν degrees of freedom and a probability level P.* Both formulae involve the use of the standardized normal deviate, y_P, corresponding to the value of P chosen.

(1) *R. A. Fisher's* (1925) *formula:*

$$\chi^2_\nu(P) = \tfrac{1}{2}\{y_P+\sqrt{(2\nu-1)}\}^2, \tag{a}$$

which assumes that $\sqrt{(2\chi^2)}$ is normally distributed about $\sqrt{(2\nu-1)}$ with unit standard deviation.

(2) *E. B. Wilson & M. M. Hilferty's* (1931) *formula:*

$$\chi^2_\nu(P) = \nu\left\{1-\frac{2}{9\nu}+y_P\sqrt{\frac{2}{9\nu}}\right\}^3, \tag{b}$$

which assumes that $(\chi^2/\nu)^{\frac{1}{3}}$ is normally distributed about $1-2/(9\nu)$ with a standard deviation of $\sqrt{(2/9\nu)}$.

Professor Pearson has suggested that I should prepare a comparative table showing for certain ν and P the numerical values of $\chi^2_\nu(P)$:

(*a*) calculated from Fisher's formula,
(*b*) calculated from Wilson & Hilferty's formula,
(*c*) the correct values taken from Miss Thompson's table (pp. 188–9 above).

* The notation used is that adopted in the paper on pp. 187–91 above.

Comparative Table of Percentage Points

ν \ P	0·995	0·990	0·950	0·900	0·750	0·500	0·250	0·100	0·050	0·010	0·005
30 (a)	13·032	14·337	18·218	20·477	24·547	29·500	34·908	40·165	43·487	50·075	52·603
(b)	13·744	14·925	18·491	20·604	24·486	29·338	34·793	40·246	43·767	50·914	53·713
(c)	13·7867	14·9535	18·4926	20·5992	24·4776	29·3360	34·7998	40·2560	43·7729	50·8922	53·6720
40 (a)	19·923	21·529	26·233	28·931	33·732	39·500	45·722	51·712	55·473	62·883	65·712
(b)	20·669	22·139	26·508	29·055	33·668	39·337	45·610	51·796	55·753	63·710	66·802
(c)	20·7065	22·1643	26·5093	29·0505	33·6603	39·3354	45·6160	51·8050	55·7585	63·6907	66·7659
50 (a)	27·188	29·059	34·487	37·570	43·016	49·500	56·439	63·072	67·219	75·353	78·447
(b)	27·957	29·685	34·763	37·693	42·949	49·336	56·328	63·159	67·501	76·172	79·523
(c)	27·9907	29·7067	34·7642	37·6886	42·9421	49·3349	56·3336	63·1671	67·5048	76·1539	79·4900
75 (a)	46·375	48·809	55·775	59·678	66·494	74·500	82·961	90·965	95·931	105·603	109·259
(b)	47·178	49·457	56·054	59·799	66·422	74·335	82·853	91·055	96·214	106·408	110·313
(c)	47·2059	49·4748	56·0540	59·7944	66·4167	74·3343	82·8582	91·062	96·211	106·393	110·286
100 (a)	66·481	69·389	77·649	82·243	90·213	99·500	109·242	118·400	124·056	135·023	139·154
(b)	67·303	70·049	77·929	82·362	90·138	99·335	109·137	118·493	124·340	135·820	140·193
(c)	67·3276	70·0648	77·9295	82·3581	90·1332	99·3341	109·141	118·498	124·342	135·807	140·169
150 (a)	108·277	111·980	122·411	128·161	138·064	149·500	161·390	172·481	179·295	192·432	197·358
(b)	109·122	112·655	122·692	128·278	137·987	149·334	161·288	172·577	179·579	193·219	198·380
(c)	—	—	—	—	—	—	—	—	—	—	—
200 (a)	151·365	155·737	167·997	174·722	186·255	199·500	213·200	225·920	233·709	248·675	254·270
(b)	152·224	156·421	168·279	174·838	186·175	199·334	213·099	226·017	233·993	249·455	255·305
(c)	—	—	—	—	—	—	—	—	—	—	—
y_P	−2·57583	−2·32635	−1·64485	−1·28155	−0·67449	0·0000	0·67449	1·28155	1·64485	2·32635	2·57583

(a) From R. A. Fisher's formula. (b) From E. B. Wilson & M. M. Hilferty's formula. (c) From Miss Thompson's table.

The comparison is shown in the preceding table.* It will be seen that:

(1) For the range of values of ν and P considered, the formula (b) is consistently more accurate than (a); the greatest absolute value of the error is about 0·04 for (b) while at the tails of the distribution it amounts to over 1·00 for (a).

(2) Formula (b) is extremely accurate in the neighbourhood of the two 5 % points ($P = 0{\cdot}950$ and $0{\cdot}050$).

(3) For neither formula do the actual error values decrease very much as ν increases but, since χ^2 increases with ν, the *relative* error values do, of course, decrease.

REFERENCES

FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

GARWOOD, F. (1936). *Biometrika*, 28, 441.

WILSON, E. B. & HILFERTY, M. M. (1931). *Proc. Nat. Acad. Sci., Wash.*, 17, 684.

* An earlier comparison of the approximations has been made by F. Garwood (1936).

Vol. XXXII. Parts III & IV April, 1942

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

Reprinted by offset-litho, 1952, 1960

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[*Issued* 22 *April* 1942]

MEDICAL STATISTICS FROM GRAUNT TO FARR

(Continued*)

BY MAJOR GREENWOOD

III. THE STATISTICAL WORK OF GRAUNT

JOHN GRAUNT's contribution to our subject has always been regarded as one of the great classics of science. A few have indeed doubted whether so great a work could have been achieved by one whose material success was so modest and have sought to transfer the glory to Graunt's highly successful friend Petty. This dispute I relegate to an appendix. I assume that Graunt's published book is substantially his own original work.

The history of the material Graunt used has been written more than once and I have nothing to add to Prof. Hull's story. Graunt had, for a period of more than 60 years, arithmetical statements of the numbers of males and females christened and buried and of the causes of death (not distinguished by sex) under some sixty headings. He had no information as to the ages at death. He had no information as to the number or ages of the living population.

The first act of a scientific statistician is to assess the trustworthiness of his data, to criticize his sources. This tedious preliminary to the doing of sums was not much to Petty's taste. Petty, as we have seen, often used different data to reach some conclusion, but hardly ever discusses the reliabilities of the several data. Other Fellows of our College since Petty's day have made the same mistake. The terrible 'howler' committed by Dr William Heberden the younger, and detected, not without satisfaction, by Charles Creighton is classical.† But that was not a unique instance. Indeed, even trained statisticians sometimes confuse names with things. More than one rate of mortality has risen (or fallen) only on paper. Graunt made no such mistakes.

Graunt's *general* argument is that many causes of death are 'but matters of sense', for instance, whether a child were abortive or stillborn, and that in many cases the searchers are 'able to report the opinion of the physician, who was with the patient as they receive the same from the friends of the defunct'. But sometimes the searchers will be wrong and often enough the error will not matter.

As for consumptions, if the searchers do but truly report (as they may) whether the dead corpse were very lean and worn away, it matters not to many of our purposes whether the diseases were exactly the same, as physicians define it in their books. Moreover, in case a man of seventy-five years old died of a cough (of which had he been free, he might

* The earlier sections were printed in *Biometrika*, **32**, 101–27.

† Creighton, *History of Epidemics in Britain*, **2**, 747–8. Heberden supposed (erroneously) that 'Griping of the Guts' of the Bills was Dysentery and had decreased. It was Infantile Diarrhoea and had simply been transferred to the rubric 'Convulsions'.

have possibly lived to ninety) I esteem it little error (as to many of our purposes) if this person be in the table of casualties, reckoned among the aged, and not placed under the title of coughs (348).*

No doubt this brutal common sense might set on edge the teeth of some Fellows of the College of Physicians even in the seventeenth century, but it was one of the qualities which made Graunt a pioneer. Making the best the enemy of the good is a sure way to hinder any statistical progress. The scientific purist, who will wait for medical statistics until they are nosologically exact, is no wiser than Horace's rustic waiting for the river to flow away.

Graunt, however, did not accept statements which he had the means of testing. Finding in a series of years that of more than a quarter of a million deaths only 392 were assigned to the Pox, he did not infer that Syphilis had been over-rated as a cause of death.

Forasmuch as by the ordinary discourse of the world it seems a great part of men have, at one time or other, had some species of this disease, I wondering why so few died of it, especially because I could not take that to be so harmless, whereof so many complained very fiercely; upon enquiry, I found that those who died of it out of the hospitals (especially that of Kingsland, and the Lock in Southwark) were returned of ulcers and sores. And in brief, I found, that all mentioned to die of the French Pox were returned by the clerks of St Giles' and St Martin's in the Fields only, in which places I understood that most of the vilest and most miserable houses of uncleanness were: from whence I concluded, that only hated persons, and such, whose very noses were eaten off were reported by the searchers to have died of this too frequent malady (356).

In principle, the argument is still valid.

His next example of criticism is the case of Rickets, which first appeared in the Bills of Mortality in 1634 and then with 14 deaths only, but by 1659 had risen to 441. Was Rickets a 'new disease' or did an old disease receive, in the Bills, a new name?

To clear this difficulty out of the bills (for I dare venture on no deeper arguments) I enquired what other casualty before the year 1634, named in the Bills, was most like the rickets; and I found, not only by pretenders to know it, but also from other Bills, that livergrown was the nearest. For in some years I find livergrown, spleen, and rickets, put all together, by reason (as I conceive) of their likeness to each other. Hereupon I added the livergrowns of the year 1634, viz. 77, to the rickets of the same year, viz. 14, making in all 91; which total, as also the number 77 itself, I compared with the livergrowns of the precedent year 1635, viz. 82. All which showed me, that the rickets was a new disease over and above. Now, this being but a faint argument, I looked both forwards and backwards, and found that in the year 1629, when no rickets appeared there were but 94 livergrowns; and in the year 1636 there were 99 livergrowns, although there were also 50 of the rickets: only this is not to be denied, that when the rickets grew very numerous (as in the year 1660, viz. 521) then there appeared not above 15 of livergrown. In the year 1659 were 441 rickets and 8 livergrown; in the year 1658 were 476 rickets and 51 livergrown. Now though it be granted that these diseases were confounded in the judgment of the nurses, yet it is most certain that the livergrown did never but once, viz. anno 1630, exceed 100; whereas anno

* Numbers in brackets are page references to Prof. Hull's edition of *The Economic Writings of Sir William Petty together with the Observations upon the Bills of Mortality more probably by Captain John Graunt*, Cambridge, 1899.

1660, livergrown and rickets were 536. It is also to be observed, that the rickets were never more numerous than now, and that they are still increasing; for anno 1649, there were but 190, next year 260, next after that 329 and so forwards, with some little starting backwards in some years, until the year 1660, which produced the greatest of all (357–8).

This is an excellent statistical argument, and, incidentally, evidence that Graunt wrote his own book, for a physician would probably have suggested that the professional interest excited by the classical treatise of Glisson (assisted by Regimonter) which was published in 1650 might easily have increased the popularity of the diagnosis. Petty, who, with Glisson, was a founder of the Royal Society, would hardly have ignored his colleague's work.

I cannot resist the desire to mention others which, while of little statistical importance, have a medical attraction. Graunt noticed that Stopping of the Stomach first appeared in the Bills of 1636, increased from 6 to 29 by 1647, by 1655 it reached 145, in 1657, 277 and 1660, 314. First he conjectured that Stopping of the Stomach might be the Green Sickness, 'forasmuch as I find few or none to have been returned upon that account, although many be visibly stained with it'. He thought that possibly Green Sickness might not appear in the Bills 'for since the world believes that marriage cures it, it may seem indeed a shame, that any maid should die uncured, when there are more males than females, that is, an overplus of husbands to all that can be wives'. Then he wondered whether Stopping of the Stomach might not be Mother, 'forasmuch I have heard of many troubled with Mother Fits (as they call them) although few returned to have died of them'. But he was diverted by guessing 'rather the Rising of the Lights might be it'. He remembered that some women troubled with the Mother fits did complain of a choking in their throats. 'Now, as I understand, it is more conceivable that the Lights or Lungs (which I have heard called the bellows of the body) not blowing, that is, neither venting out, nor taking in breath, might rather cause such a choking, than that the Mother should rise up thither, and do it. For methinks, when a woman is with child, there is a greater rising, and yet no such fits at all' (359). He notes that Rising of the Lights increased in the Bills from 44 in 1629 to 249 in 1660.

Finally, he suggests a correlation between Stopping of the Stomach, Rising of the Lights in adults and the Livergrown, Spleen and Rickets of children. 'And that what is the Rickets in children, may be the other in more grown bodies; for surely children which recover of the Rickets, may retain somewhat to cause what I have imagined: but of this let the learned physicians consider, as I presume they have' (359).

It might be suggested that one item under Stopping of the Stomach could be surgical, viz. strangulated hernia. Rupture was a heading in the Bills, but the numbers are small and show no regular increase with the increase of population. Graunt's attraction to what used to be called hysterical stigmata is interesting. One wonders how far these passages reflect conversations with Petty. It is clear that Graunt had no belief in the peripatetic uterus; Petty

would have had none. The best medical opinion of the age is, of course, that of Sydenham. Sydenham (whose pathology was traditional) had a pneumatist aetiology of Hysteria, the origin was an ataxia of the animal spirit (which was the pneuma zotikon of ancient tradition). He not only believed that Hysteria might be a serious or even mortal complication of organic disease—as we do still—but that the ataxic spirits might themselves produce humoral corruption and lead to chlorosis or ovarian dropsy (*Dissertatio epistolaris*, 92). So there is nothing repugnant to the best professional opinion of the age in admitting Hysteria to the list of causes of death. Nor is there any gross absurdity in the suggested correlation of increasing Rickets and increasing Hysteria, from the point of view of a layman. But that surmise does not imply any professional hint, it rather suggests a belief in a merely physical factor, the pressure of an enlarged organ. That passage would not have been written by a physician.

These are sufficient instances of Graunt's criticism of sources—the temptation to go on quoting examples must be resisted. I pass to his great achievement, the estimation of rates of mortality at ages when the numbers and ages of the living were not recorded. For such an estimation to be correct, we all know that the population must be stationary, viz. non-increasing, not subject to migration and having constant rates of mortality in the several age groups.

It is a nice point whether Graunt or Petty appreciated the importance of these considerations. Graunt was certainly alive to the fact that the population of London was growing and that the growth was due to immigration from the country. The arithmetical position was this. In the earlier years of his series burials and christenings were about equal in numbers, in 1605 there were 5948 burials and 6504 christenings; in 1625, 7850 burials and 7682 christenings, in 1635, 10,651 burials and 10,034 christenings. Later the burials continued to increase, but the christenings either decreased or failed to increase in the same proportion. This Graunt attributed to neglect of christening owing to religious dissidence and gave excellent reasons for his view. It is clear then that there were two factors of increase, immigration and increasing numbers of births. Most of Graunt's deductions are based upon an analysis of the deaths by causes for twenty years, 1629–36 and 1647–58, which he selected as years comparatively unaffected by plague (of his total of 229,250 deaths only 16,000 were from plague).

If we treat this total as a denominator (or one-twentieth of it) it will, from the point of view of calculating mortality ratios, be affected by two errors. The deaths of immigrants will make it too large and the increasing births will make it too small. Can it be Graunt held that the errors balanced so that, arithmetically speaking, one might behave as if one were dealing with a stationary population? An alternative explanation is that Graunt did not realize the limitations of the method.

A third possibility is that, although he knew the fallacy, he believed that

the incorrect method gave an approximation to truth sufficient for his purposes. This is the solution I should be inclined to adopt were I forced to choose.

As I have pointed out above, there is at least a suggestion that Petty did have some glimmering of the conditions to be fulfilled if a summation of deaths is to give a correct view of rates of mortality. I do not believe that Graunt was less informed on any point of vital statistics than Petty. However, all this is guess-work.

Graunt did not know the ages of the dead; what he did was to pick out of the list of causes of deaths those which he thought lighted only upon children 'not more than four or five years old'. He chose Thrush, Convulsions, Rickets, Teeth and Worms, Abortives, Chrysomes, Infants, Livergrown and Overlaid. These gave him some 70,000 out of some 229,000. Then he assigned half the deaths from Small Pox, Swine Pox, Measles and Worms without Convulsions also to children under six and reaches the final conclusion that about '36 % of all quick conceptions die before six years old'.

Is this conclusion—I will not say correct, because we have no data to reach a correct result—but of a reasonable order of magnitude? The answer is that it is eminently reasonable. Two hundred years after Graunt's death, William Farr printed (in the famous Supplement to the *35th Annual Report of the Registrar-General*, p. cxxxvi) an outline Life Table for London. This was, of course, computed by an approximately correct method, using knowledge of the numbers and ages of the living population, and reflects the conditions of seventy-five years ago. Interpolating in this we find that about 32 % of 'quick conceptions died before six years old'. There is no good medical reason for holding that the conditions of *child* life in London in the middle of Victoria's reign were much better than in the seventeenth century. The old genius used a bow with a frayed string and made no allowance for windage, but his arrow hit the target not far from the white. He gave the first quantitative measure of the Herodian sacrifice in towns, a sacrifice which was to continue to be offered for more than 200 years.

Graunt then passed to the other end of life and found that 7 % of the dead were 'aged'. He conceived that the searchers would mean by 'aged' persons of 70 years or upwards, 'for no man be said to die properly of Age who is much less'. His following suggestion that the proportion living beyond 70 might be used as a measure of healthfulness is not happy. But this calculation may have led him to make, or insert, the most famous passage in his book, viz. what is, in form, the first Life Table ever published.

> Whereas we have found, that of 100 quick Conceptions about 36 of them die before they be six years old, and that perhaps but one surviveth 76; we having seven decads between six and 76, we sought six mean proportional numbers between 64, the remainder, living at six years, and the one, which survives 76, and find that the numbers following are practically near enough to the truth; for men do not die in exact proportions, nor in fractions, from whence arise this Table following (386).

Graunt's figures are 100, 64, 40, 25, 16, 10, 6, 3, 1.

The one survivor to 76 is, as Graunt implies, a guess; perhaps he conjectured that his seven survivors beyond 70 died one a year. How he calculated his mean proportional numbers is unknown. Prof. Willcox conjectured that he experimented with multipliers of 5/8 and 2/3—the former nearly reproduces the figures (see Willcox, *Revue de l'Inst. Intern. de Statistique,* 5 (1937), 327). Ptoukha (*Congrès Intern. de la Population*; Démographie historique, p. 71, Paris, 1937) ingeniously suggests that he used the multiplier $(64-1)/100$ or 0·63.

We must, I fear, conclude sorrowfully that this shot did not find the bull's eye. If Graunt's survivors are compared with those shown in Halley's table (when correctly used, *vide infra*), for 100, 64, 40, 25, 16, 10, 6, 3, 1, we should have 100, 56, 50, 45, 38, 31, 22, 14, 6. It is possible that child mortality was lower in London than in Breslau, but quite incredible that later age mortality should have been so enormously higher.

But, of course, having regard to the data, it would have been more than genius, it would have been magic, had a correct result been obtained.

Prof. Willcox, whose opinion of Graunt is almost as high as mine, regards the passage as inserted on the recommendation of Petty and as Petty's composition. He thinks that it lacks Graunt's caution and suggests the flighty ingenuity of his friend. Prof. Willcox's arguments are weighty, but I am not convinced. That Graunt did not—to use the expressive slang—*feature* his table is true. It is also true (*vide supra*) that passages in Petty's undoubted writings imply that he had some conception of a survivorship table. But—and this is my main difficulty—if this were Petty's idea, I find it difficult to believe that he would not have exploited it. Halley, whose economic scent was not so keen as Petty's, saw the epoch-making importance of an idea which was to transform the business of selling annuities. It would be odd if Petty *had* seen it that he did not comment upon it. Graunt might well have hesitated, being a cautious statistician, but surely not Petty.

However, in spite of modern practice, the writing of history wholly in terms of psychology has its pitfalls.

Let us return to simpler applications of shop arithmetic. The advantages of country life over town life from the point of view of both mortality and morality had been a commonplace of poets, particularly those Roman poets who spent much of their lives in a city, long before the seventeenth century. Graunt was the first to apply an arithmetical test of mortality; he compared the statistics of Romsey with those of London. For Romsey he had ninety years' data of marriage, christenings and burials.

His statements about the population of the parish are not quite consistent. In one sentence he says that it 'both 90 years ago, and also now, consisted of about 2700', but a few lines later says 'it neither appears by the burials, christenings, or by the built of new housing, that the said parish is more populous now, than 90 years ago, by above two or 300 souls'. A little later he says 'it is

clear that the said parish is increased about 300, and it is probable that 3 or four hundred more went to London; and it is known that about 400 went to New England, the Caribe Islands and Newfoundland within these last forty years' (389). Actually, from an estimate of the number of communicants (which he assumes to be rather more than half the total population) he makes the average population between 2700 and 2800. Taking the average of burials for the whole period to be 58, this gives him a death rate of a little more than one in 50, which he contrasts with the London figure of one in 32 (apparently based on his count of 11 families with 88 persons amongst whom 3 deaths occurred in a year; but this is a rate of one in 29).

There is no doubt a certain sketchiness about this, but it was not unreasonable to infer that the Romsey rate was much lower than the London rate.

Graunt found that, unlike London, Romsey had an average excess of christenings over burials, they were in the ratio of 5 to 4. He estimates that over the period the natural increase was 1059, and, as will be seen from the quotation made, he allots about a third of this respectively to London, to the colonies and to the parish itself. He argues that supposing the population of all England to be fourteen times that of London and other parishes to send one-third of their natural increase to London, then the London burials should increase about 200 per annum 'and will answer the increase we observe'.

Here again the argument is reasonable. He goes on to an investigation which has been severely criticized. He gives a table of the greatest and least number of burials in each of the ten-year periods for which he has data. In each decade but one the maximum is more than twice the minimum. But, he remarks, in no decade in the London experience is the largest number of burials twice the smallest number (he excludes deaths from plague from his statistics). 'Which shews, that the opener and freer airs are most subject both to the good and bad impressions, and that the fumes, steams and stenches of London do so medicate and impregnate the air about it, that it becomes capable of little more, as if the said fumes rising out of London met with, opposed and jostled backwards the influences falling from above, or resisted the incursion of the country airs' (392).

Prof. Hull shook his head over this passage. 'This is an attempt to explain by physical conditions the wide range in the observed country death rate which is really due to the narrowness of the field—a single market town—under investigation. It is perhaps the gravest statistical mistake that can be charged against Graunt' (lxxvii).

I do not like to leave a hero in the lurch. I must concede that if both Romsey and London burials were samples from a Poisson universe, the fact that the Poisson parameter for London was at least a hundred times that for Romsey would make it incredible that the London range, in terms of the mean or of the standard deviation, should be so wide as that for Romsey. But Prof. Hull was

wrong in supposing that the wide range in the Romsey rates was due to the narrowness of the field of observation in a statistical sense.

Taking Graunt's 58 as the 'expected' annual deaths then, as 1/58 is small, the Poisson distribution is not far from the symmetry of a normal curve, and using the results of Tippett and E. S. Pearson, we may conclude that the expected range would be $23{\cdot}45 \pm 6{\cdot}073$. The observed ranges for the successive decades are 32, 48, 78, 23, 65, 39, 121, 91, 52. All but one is greater than the expectation and six diverge by more than three times the standard error.

Something more than small numbers is involved. Still, it must be confessed that Graunt did not anticipate the reasoning of James Bernoulli, although an intuition of genius may have led him to think that something more than 'chance' had play here.

Graunt devoted special attention to the demographic influence of the plague. In the first place, he remarks that the attribution plague understated the mortality due to plague. He infers this from the fact that in plague years burials from other causes exceeded the average greatly, 'from whence we may probably suspect, that about 1/4 part more died of the plague than are returned for such'. Next he inferred that after a great outburst plague lingered for several years.

> The plague of 1636 lasted twelve years, in eight whereof there died 2000 per annum one with another, and never under 300. The which shows that the contagion of the plague depends more upon the disposition of the air than upon the effluvia from the bodies of men. Which also we prove by the sudden jumps which the plague hath made, leaping in one week from 118 to 927; and back again from 993 to 258; and from thence again the very next week to 852. The which effects must surely be rather attributed to change of the air, than of the constitutions of men's bodies, otherwise than as this depends upon that (366).

Finally, he observes that within two years the city was re-peopled; a deduction from the time taken for the number of christenings to reach again the level of a pre-plague year.

We may, if we please, smile at Graunt's epidemiological inference. But it is a reasonable inference from the facts when we remember that in Graunt's day—in spite of Fracastorius—contagium was not thought of as contagium vivum, but as a mere sympathetic vibration or passing on of something.

I have, I hope, given an adequate sample of Graunt's quality, but have not mentioned the most famous of all his deductions. Both in London and the country, on the average more males were christened than females, but more males died young or entered celibate occupations. So we reach this conclusion:

> We have hitherto said, there are more males than females; we say next that the one exceed the other by about a thirteenth part. So that although more men die violent deaths than women, that is, more are slain in wars, killed by mischance, drowned at sea and die by the hand of justice; moreover more men go to colonies and travel into foreign parts than women; and lastly, more remain unmarried than of women as fellows of colleges, and apprentices above eighteen etc. yet the said thirteenth part difference bringeth the business but to such a pass, that every woman may have an husband, without the allowance of polygamy (375).

The story of how this arithmetical justification of God's providence attracted the attention of Derham, of how Derham's book fired the enthusiasm of the Prussian Army chaplain Johann Peter Süssmilch and of how Süssmilch's book influenced Malthus has been well told by Hull. I should not myself rank this section high among Graunt's researches. From a demographic point of view neither judicial hangings nor college fellowships could have had much effect in reducing the male excess.

Even copious quotation fails to convey the spirit of a complete book. I have quoted good things, but many more remain. Graunt revealed sundry important truths and not the least important was that very imperfect data, if patiently considered, will tell us something it is good for us to know. If young medical officers going to parts of the empire where organized medical and demographical information is at no higher a level than that of seventeenth-century England—and there are many such places—were restricted to a single book on statistics, I should advise them to take not a modern scientific work, but old John Graunt's *Observations*.

APPENDIX

Did Graunt write the book published over his name?

John Graunt and William Petty were, as we have seen, close friends. Graunt, as the world judges success, failed and Petty succeeded. But by the judgment of scientific men in the seventeenth century, and ever since, the order of intellectual precedence was reversed. From the moment of publication a few discerning people perceived the originality and importance of the *Observations*, the same people who, while admiring Petty's verve, ingenuity and worldly success, did not take over-seriously his bright ideas.

But Graunt was a man of one book. Save a note upon the multiplication of carp and the growth of salmon, he published nothing more. Petty went on writing, scheming and talking for thirteen years after Graunt's death. That often enough in that period, the *Observations* were discussed over the wine—as they were quoted in Petty's writings—we may suppose. That Graunt discussed his work with Petty both before and after publication we may also take for certain, although we have no formal proof of it. The country statistics which Graunt first used were from Petty's native parish and even if we are not disposed—as certainly I am not—to give much weight to particular turns of phraseology, still there are sufficient verbal oddities in some pages of Graunt's book to suggest Petty's hand.

In these circumstances, it would not be very surprising if Petty's associates, particularly those who were not good judges of statistical work, were to conclude that Petty's share in the remarkable achievement of Graunt were greater than

appeared. It is not even judging Petty too harshly to suppose that he himself might come to share the opinion. There is no evidence that Petty ever *did* explicitly claim the credit. In one list of his writings (one of four), found among Petty's Papers, he did include the *Observations*, which at least is evidence that he thought himself entitled to a share of the credit. We may suppose that if, in familiar intercourse, somebody had said 'Come, confess Sir William, yours was the hand that guided the pen of poor John Graunt', he might not have denied it very strenuously. I think I have produced evidence enough that Petty did not under-rate his powers and was not conspicuous for delicacy of feeling. My guess would be that long before his death he did come to believe that Graunt's intellectual success was due to his help.

Whether Petty believed this or not, it is certain that friends and associates of Petty began to believe it soon after Graunt's death, and the belief has been entertained by a few people in each generation since. These, with one conspicuous exception, have been drawn from Petty's friends or descendants or from literary critics.

In the seventeenth century, of Petty's circle, Evelyn, Southwell and Aubrey believed or said that Petty wrote or inspired Graunt's book. Two Fellows of the Royal Society, Houghton and Halley, also attributed the book to Petty. The only one of the five who was certainly a competent judge of scientific merit was Halley. Halley began the memoir which contains his Breslau table with these words:

> The contemplation of the mortality of mankind has, besides the moral, its physical and political uses, both which have some time since been most judiciously considered by the curious Sir William Petty, in his moral and political Observations upon the Bills of Mortality of London, owned by Captain John Graunt. And since in a like treatise on the Bills of Mortality of Dublin....But the deductions from those bills of mortality seemed even to their authors (*sic*) to be defective. (*Phil. Trans.* no. 196 (1693), p. 596.)

Since the seventeenth century, there has been unanimity among demographic statisticians and economists that Petty could not have written Graunt's book. Halley was quite as good a judge of scientific merit as any of them and a contemporary of the canvassed writers; if I were sure that he had read and compared Petty's acknowledged works with the *Observations* I should prefer his opinion to that of other 'experts'—including, of course, my own. Halley's direct testimony, in the sense of a court of law, would be valueless; he was only six years old when the *Observations* were published and became a Fellow of the Royal Society five years after the death of Graunt. There is no evidence that either before or after the period of writing and publishing his famous memoir, Halley worked on demography. After his memoir, but in his lifetime, a new epoch in mathematical vital statistics began. De Moivre, eleven years younger than Halley, brought out his principal works in the lifetime of Halley (1656–1742) and used Halley's table. The two men must have been well acquainted,

for both were enthusiastic disciples and intimate friends of Newton, but Halley, like Graunt, made only one contribution to the literature of demography.

So it may be doubted whether Halley were sufficiently interested in demographic or economic writings to have read Petty's tracts at all. Also in the passage cited above (apart from the writing of 'authors' not 'author') the collocation of the *Observations* on the Dublin Bills with those on the London Bills is curious. There is no doubt that the *Observations* on the Dublin Bills were the work of Petty, and in the first edition of them they are stated on the title-page to be by the 'Observator on the London Bills of Mortality'. But this, as Prof. Hull pointed out (xlii), was probably a catch-penny device of the publisher, Mark Pardoe, to draw a public which had just taken a fifth edition of the London *Observations*. Actually the book did not sell, and when the publisher reissued an enlarged version, Petty's name appeared on the title-page without any reference to the London *Observations*. I conclude that Halley's evidence is less weighty than it seemed. He may well have had before him copies of Graunt's book and of the two editions of the Dublin *Observations*. Having no other knowledge of the literature he would naturally enough write as he did.

If we eliminate Halley, no other expert countenanced Petty's authorship and one, Augustus de Morgan, gave an amusing but quite cogent reason for dismissing the notion.

In speaking of the variations in the annual numbers of deaths attributed to Rickets, Graunt said:

> Now, such back-starting seem to be universal in all things: for we do not only see in the progressive motion of wheels of watches, and in the rowing of boats, that there is a little starting or jerking backwards between every step forwards, but also (if I am not much deceived) there appeared the like in the motion of the moon, which in the long telescopes at Gresham College one may sensibly discern (358).

De Morgan (*Budget of Paradoxes*, 68; *Assurance Magazine*, 8, 167) commented on the improbability that 'that excellent machinist, Sir William Petty, who passed his day among the astronomers' would attribute to the motion of the moon in her orbit all the tremors which she gets from a shaky telescope.

Down to 1927 the matter was regarded, in scientific circles, as settled. In that year the late Marquis of Lansdowne published a copious selection of the Petty Papers with what he regarded as new evidence in favour of Petty.

The only new evidence of a direct kind was a manuscript list in Petty's hand of his writings or projected writings which included the *Observations*. There are three other lists which do not include the *Observations*, and if we are to suppose that the entry really referred to the book published under Graunt's name, then we must believe that in 1685 and in 1686 Petty had forgotten his best title to scientific immortality. The remainder of the evidence consists of parallel passages and *ad captandum* arguments to the effect that it was more probable that a physician had written on questions of medical statistics than a tradesman. This

publication led to a lively controversy. Of the merits of this, I, as a party to it, am not an impartial judge. Purely literary arguments do not appeal to me when the question is of scientific method. Thus, Dr L. F. Powell attached weight to the fact that Dr Johnson in conversation had attributed to Petty an observation (not statistical) which is made not in Petty's writings but in Graunt's book.

In the discussion the word 'style' is used in different senses by the combatants. The statisticians are thinking of scientific method, the literary critics of verbal arrangement. To the former the fact that, particularly in the conclusions and the Appendix, Graunt's book has turns of phraseology which suggest Petty's hand, seems of little importance. To the latter it seems very significant.

In the article by Prof. Willcox, which I have quoted above, the controversy is reviewed, and the author concurs generally with his statistical predecessors.

Prof. Willcox does, however, differ from his predecessors in one important particular. He holds that the famous life table was supplied by Petty. He argues that this is far too conjectural to have been the work of so cautious a reasoner as Graunt:

> In attempting to reconstruct its origin I have surmised that after Graunt had estimated that 36 per cent. of the deaths were due to children's diseases, that they all occurred under the age of six, and that the seven per cent. who were reported to have died 'aged' died at over 70 years of age (at one place he says over sixty), he felt unable to go further and reported his difficulty to Petty, already perhaps speculating about a series of similar problems.
>
> Petty guessed at the number of survivors at the end of each decennial age period, 6–15, 16–25 etc. incidentally and characteristically ignoring Graunt's theory that seven per cent. survived seventy, and assuming instead, without reason, that one per cent. survived seventy-six and not one per cent. eighty-six, and that the survivors at age six decreased with each age period in a geometrical progression approximately equal to the 64 per cent. which Graunt had set for the first group (326–7).

Prof. Willcox's argument is cogent. It may be strengthened by a criticism of the late Prof. Westergaard (*Contributions to the History of Statistics* (London, 1932), p. 23). In using this table, Graunt made a serious blunder. In order to estimate the number of men of military age in London he subtracted the number alive at age 56 from the number alive at age 16. But this simply gives him the number *dying* between those ages; what he wanted was some average of the l_x's. It is evident that Graunt was not at all clear in his mind as to how to use a life table.

On the other hand, if this table were really Petty's idea, it is hard to understand why he did not exploit it. If Petty had been a Halley, the explanation would be obvious. The table is wrong; the conditions for the validity of the method were not fulfilled. There is indeed (*vide supra*) some evidence that Petty did know what data were necessary in order to construct a proper life table. One seems on the horns of a dilemma. If Petty thought the table was correct why did he make no further use of the method? If he thought it was wrong, would he have urged Graunt to insert it?

Although Prof. Willcox has certainly shaken my previous conviction, I still feel reluctant to surrender Graunt's table to Petty. However, there may be an element of sentimentality in this. At least the statisticians agree that the answer to the question which I have placed at the head of this Appendix is emphatically *yes*.

REFERENCES TO THE RECENT CONTROVERSY IN CHRONOLOGICAL ORDER

1927. LANSDOWNE, Marquis of. *The Petty Papers.*
1928. GREENWOOD, M. *J.R. Statist. Soc.* 91, 79.
1928. LANSDOWNE, Marquis of. *The Petty Southwell Correspondence*, pp. xxiii–xxxii.
1932. LANSDOWNE, Marquis of. *The Times Literary Supplement*, 8 Sept.
1932. BRETT-JAMES, N. G. *The Times Literary Supplement*, 15 Sept.
1932. GREENWOOD, M. *The Times Literary Supplement*, 22 Sept.
1932. LANSDOWNE, Marquis of. *The Times Literary Supplement*, 13 Oct.
1932. POWELL, L. F. *The Times Literary Supplement*, 20 Oct.
1933. GREENWOOD, M. *J. R. Statist. Soc.* 96, 76.
1937. WILLCOX, W. F. *Revue de l'Inst. Intern. de Statistique*, 5, 321.

IV. HALLEY'S LIFE TABLE

The long and fruitful life of Edmund Halley (1656–1742) belongs to the general history of science; of him it may indeed be said *nihil quod tetigit non ornavit*. He made only one contribution to our subject, but it was of first-rate importance.

The circumstances of this undertaking are obscure; Halley would have perceived the imperfections of Graunt's life table, but it is not known whether it was he who set on foot a search for better statistical material than Graunt had had. Inquiries were however made, and made after he had become a Fellow of the Royal Society, so it is at least possible that Halley, who had travelled extensively in Europe (he was at Danzig in 1679 and in Italy in 1681), suggested that something might be found abroad. By 1691, the King's Librarian, Henry Justel, who was in touch with the Society, had been brought into communication, possibly through Leibniz, with Caspar Neumann, a scientifically minded evangelical pastor of Breslau. Neumann supplied the data which Halley used.

In 1883, J. Graetzer, a medical-statistical official of Breslau, published a little monograph* which throws light upon the work. He not only extracted from the Breslau archives all the data which were or might have been communicated to the Royal Society but had the Society's archives searched, with the result that a letter from Neumann to Justel and another from Neumann to Halley, both with statistical appendices, were discovered. Thanks to the labours of Graetzer and an essay by R. Böckh (*Bulletin de l'Inst. Intern. de Statistique*, 7 (1893), 1) we can form a reasonably clear idea of Halley's method, which was not what those who have not examined the literature suppose it to have been.

It is often stated that Halley, having found that during the five years of observation the number of births only slightly exceeded the number of burials,

* Graetzer, J., *Edmund Haley und Caspar Neumann*, 1883.

treated the population as stationary and constructed a life table by a simple summation of the deaths in the manner already explained. He was much wiser. What he tried to do was to construct a *population* table, in the following way. Suppose we know how many children were born in a calendar year, say 1690, in a town not subject to migration which maintains accurate registers of ages at death, and then we discover how many of the children born in 1690 will be alive on each successive first of January by a series of subtractions. We shall have the survivors on 1 January 1691 by subtracting those of the children born in 1690 who died in 1690. We shall have the survivors on 1 January 1692 by subtracting the deaths in 1691 occurring among the survivors to 1 January 1691, and so on. This will give a precise enumeration of the living population, and this is what Halley wanted. The figures we shall obtain will *not* be the conventional l_x's of a Life or (as German writers say) Mortality Table, but what in most modern books are represented by the capital letter L or years of life lived or "persons" living between the termini (see Appendix). *If* the population is stationary, the sum of these figures gives the population of the place under study. Now for ages between 1 (last birthday) and very advanced ages, L_x is simply l_x diminished by $\frac{1}{2}(l_x - l_{x+1})$. In the first year of life (and at advanced ages) the difference is greater. Thus in the first year of life deaths are not evenly distributed throughout the year of life, more than 70% of them occur in the first six months of life, so that instead of subtracting half the deaths we must subtract nearly three-quarters. Halley himself assigned 68% of deaths in the first year of life to the first half of the first year. The reason why Halley proceeded in this way was that he knew the population *not* to be stationary. His idea was to obtain the figures for the first few years of life accurately—indeed just as they are now obtained—and then to correct for excess of births over deaths.

His masterly plan was partly defeated by the fact that his Breslau correspondent Neumann was not so good a statistician as he was. Halley's letter to Neumann has not been preserved, we have only Neumann's answer of 1 March 1694. Probably Halley asked Neumann to send him (as a check on the calculations he had already made) the exact numbers of survivors on 1 January for five years, of births in a calendar year. Neumann did send him a table, but the table, as Graetzer pointed out, is wrong. Neumann gave the correct figures for 1 January of the first successive year, incorrect figures for the other years. Between 1 January of the year following the year the births of which are under study, and the next first of January, some of those born in the starting year will die under and some over a year of age. Neumann merely deducted the former, so he has too many survivors. To reach the right figure would have meant taking more trouble and he did not appreciate the importance of this. Böckh—whose opinion of his statistical contemporaries has a tinge of bitterness rare, of course, in other scientific pursuits—remarks that it was not strange Neumann should miss the point as it had been missed by many statisticians long after his time.

s at least clear that Halley had realized an important truth which did not ome part of even expert knowledge for more than a century.

The precise arithmetical details of Halley's work are not perhaps of much lical interest. Graetzer and Böckh have done a good deal to clear it up. The a used (an average of five years) had 1238 births and 1174 deaths and the le accounts for 1238 deaths. Halley must therefore have had a plan for easing deaths. It is likely, from an observation he makes on mortality in ist's Hospital, that he did not *wholly* depend on the Breslau figures. Graetzer gests that Halley may have made two graphs, one having an ordinate of 1238 he origin and an ordinate of 64 at the oldest age, the other an ordinate of 4 at the origin and 0 at the oldest age and that he plotted the survivors for h graph based on recorded deaths and drew a curve passing through 1238 l 0 between these graphs. It may be so. Using the original material which etzer published, Böckh recalculated the table. The results do not, except at s over 60, differ materially from Halley's. So far as concerns the mean after time (expectation of life), Halley's table gives 27·54 years at birth, Graetzer's terial 27·69. For ages under 40 the re-working gives slightly lower and for r ages higher mortality. It may be noted that Halley's table gives appre-oly *higher* mortality in childhood than Graunt's, more than 43 % instead of % are dead by the age of six years. But Graunt's method would exaggerate rtality (so would Halley's method, but, owing to his precautions, not so atly). On the other hand, Graunt's estimate of age is only an intelligent ss. Actually, as Graetzer showed, the infant and child mortality shown by lley's table differed little from the observed rates of mortality in the city of slau in 1876–80.

It has been said that Halley was not greatly interested in the medical aspects iis work. After describing methods of calculating the prices of annuities, he the following passage*:

> It may be objected that the different salubrity of places does hinder this proposal from ig universal; nor can it be denied. But by the number that die being 1,174 per annum 34,000 it does appear that about a 30th. part die yearly as Sir William Petty has iputed for London; and the number that die in infancy is a good argument that the air it indifferently salubrious. So that by what I can learn, there cannot perhaps be one er place proposed for a standard. At least 'tis desired, that in imitation hereof the ous in other cities would attempt something of the same nature, than which nothing iaps can be more useful.

That the mortality of childhood depends upon the atmosphere is not so lish an hypothesis as it may seem to us. Halley lived before breast-feeding ame the exception rather than the rule. The 'curious' in other cities had not wit to follow his advice. He made no other contribution to the science of il statistics; a gain to astronomy but a heavy loss to demography.

* I have read Halley's paper in the collection of papers, many by him, collected under title *Miscellanea Curiosa*, printed in London in 1705; the quotation is from p. 300.

Appendix

Halley's table is printed in two columns, the first headed 'Age current', the second 'Persons'. Thus:

Age current	Persons
1	1000
2	855
3	798
4	760
5	732
6	710
7	692
8	680
9	670
10	661

and so on.

A mistake sometimes made is to suppose that Halley meant by age current simply the end of each year of life and that the entry against each 'age current' is the number of survivors at exact age one year less than the 'age current', viz. that of 1000 born 855 survived to the first anniversary, 798 to the second anniversary, etc. The fact that Halley uses the round number 1000 for a first entry does something to encourage the mistake among readers who have not consulted the original paper and it is sometimes made by people who should know better. It is actually a terrible 'howler', leading to a wholly false view of rates of mortality in early life. Thus if 1000 and 855 were really the first two entries of a Life Table as set out now, then, as the first two entries in English Life Table no. 7 Males (mortality 1901–10) are 1000 and 856, we might conclude that mortality in the first year of life was no lower in 1901–10 than in Breslau in the last years of the seventeenth century. But the 1000 of Halley's table is *not* the number of new-born children but the average number out of 1238 born living between the ages of 0 and 1. This is what is called the L_x of a modern table or the population living between the ages x and $x+1$. If we have a column of L_x's, which is what Halley gives us, we can deduce therefrom the more familiar l_x's provided we know the starting value and the number of deaths in the first year of life. Halley gives both items. He says that of 1238 annual births 348 die annually. So that his l_0 is not 1000 but 1238, and his l_1 is 890. He chose 1000 for L_0 by assuming that of the 348 deaths in the first year of life 238 occurred in the first six months of life, 68%. This differs very little from the modern practice; in Life Table no. 7 quoted above 73·5% of the deaths in the first year of life are assigned to the first six months of life. Having been given l_0 and l_1 we can deduce the other l's from the values of the L's which Halley gives because, after the end of the first year of life there is little error in supposing that the deaths between two birthdays are evenly distributed over the year; so, for instance, l_2 will be equal to L_1 less half the difference between l_1 and l_2,

and proceeding in this way we put Halley's table into modern form. I attach a table calculated by Böckh.

It will be seen that, if Halley's table is properly used, the comparison is not of 1000 and 855 with 1000 and 856 but of 1000 and 719 with 1000 and 856.

Actually this is still slightly optimistic, because I am comparing 'persons' with males. The 'persons' figure for 1901–10 is 1000, 869. On the other hand, in the Breslau data still births (or some of them) are included in births, so that the mortality is slightly exaggerated. If for instance 7% are still born, the survivors to 1 will be the same, but the l_0 should be reduced to 930. Or alternatively we should write 1000 and 773.

I attach Halley's table reduced to modern form and with the corresponding expectations of life calculated by Böckh (I have reworked some of the values from the data and agree with Böckh's figures).

Halley's Table, expressed in modern form together with the Expectations of Life at quinquennial intervals (Böckh)

Age	l_x	$\mathring{e}_x$	Age	l_x	$\mathring{e}_x$
0	10,000	27·54	40	3,557	22·05
5	5,816	41·47	45	3,167	19·47
10	5,307	40·25	50	2,751	17·05
15	5,049	37·19	55	2,319	14·75
20	4,806	33·93	60	1,914	12·33
25	4,552	30·69	65	1,511	9·96
30	4,257	27·64	70	1,103	7·74
35	3,921	24·78	75	670	7·50

V. GUESSING THE POPULATION

My object is to trace the growth in our country of that part of statistical science which is of interest to students of medicine or public health. In speaking of such pioneers as Graunt, Petty and Halley it was proper to construe the obligation rather freely. Both Graunt and Petty did clearly perceive the relevance of their researches to matters of public health or even clinical medicine, but much of what Petty did had a more direct bearing upon political questions than those of public health. Again, the life table is a way of expressing the facts of mortality which is valuable in some medical researches, but its importance as a statistical instrument has been much greater in non-medical than medical circles, above all of course in the financing of assurance business. The commercial importance of life tables was perceived by Halley and by other mathematicians of his and the following generations.

Looking at the position after Halley's publication it was clear that progress might be made (1) in improving the accuracy of the life table, viz. by obtaining data more relevant to the conditions of life of persons who assured their lives or bought annuities, (2) in simplifying the very laborious calculations which the

determination of praemia or purchase values required. Under (1) no progress worth speaking of was made in England until the end of the eighteenth century. This was partly due, as we shall see, to a not entirely unjustified disbelief in the powers of the medical profession to change the rate of mortality, partly to ignorance. No first-rate English mathematician after Halley gave any critical attention to the theory of the Life Table before the nineteenth century. Under (2) considerable progress was made, but this progress is of little or no medical interest and to describe it would involve entering upon tedious arithmetical and algebraical detail. The primary medical-statistical quaesita are correct enumerations of deaths by sex at ages and by causes, and of the numbers living in sex and age groups. When these have been satisfied, the medical statistician can get to work.

For 150 years after Graunt's death very little was done to improve matters. Down to 1801 the population as a whole had not been counted; forty years more passed before a reasonable age distribution was secured, and it was thirty-eight years after the first denominator (populations) that the first numerator (deaths) of the fundamental fractions was obtained. Until 1801 intelligent guessing was the method and the guesses of the eighteenth century deserve a few pages, if only because they prove that statistical ability is as rare as other kinds of ability and that wishful thinking is not a modern foible.

The first estimator to mention belongs to the seventeenth century and was a younger contemporary of Graunt and Petty, Gregory King (1648–1712). He was born in Lichfield, the son of a land surveyor. At the age of fourteen he was recommended as a clerk to the famous herald Dugdale with whom he worked for several years; after Dugdale had finished his Visitation, King worked for various amateur antiquaries and was eventually invited by a lady of property in Sandon (Staffordshire) to be her steward, auditor and secretary. Here he remained until 1672 when he moved to London and, no doubt through Dugdale's recommendation, had a considerable amount of employment in both heraldic work and ordinary surveying. In 1677 he became a member of the College of Arms, in which he attained the rank of Lancaster Herald and so continued until his death, but worked for other official bodies on financial subjects.

The decorous memoir by George Chalmers, from which I have extracted these particulars, does not give us a very life-like picture of the man himself. There is a certain likeness between the early careers of Petty and King. King was not indeed shipped as a cabin boy, but Mr King (the elder) drank (if we may venture so coarse an abbreviation of Chalmers's statement that the father studied and practised his profession 'with more attention to good fellowship than mathematical studies generally allow') and King junior was a pupil teacher at eleven. If he really read Hesiod and Homer, made Greek verses and taught himself to survey land in his thirteenth year he must have had Petty's precocity. Both Petty and King had experience of practical surveying and, of course, both

were interested in political arithmetic. But there the parallel ends. King was a professional surveyor and archivist and had a reasonably successful professional career. Petty was—Petty. One might, perhaps, adduce as another parallel that King made some enemies and thought himself ill-used. But the job by which Sir John Vanbrugh, a stranger to the College of Arms, was made a king-at-arms over the head of an official of twenty years' standing would have galled the meekest of mankind. One may safely conclude that King had more knowledge of the data of political arithmetic than Petty and less originality. His vital statistical work was not published until nearly a century after his death, as an appendix to the second edition of *An Estimate of the Comparative Strength of Great Britain* by George Chalmers, London, 1803. Perhaps he never intended to publish it—he communicated the substance to his contemporary Davenant—and this may explain why there are no details of how some of his results were reached. The report reads rather like a document prepared for official use by persons interested in results not methods.

The starting-point of King's attempt to estimate the population was a return from the Hearth Office of the number of houses assessed to tax on Lady Day, 1690. That was 1,319,215 which, King estimated, had increased to 1,326,000 by 1695. He deducted 30,000 for empty divided houses,* took the round figure of 1,300,000 and assigned 105,000 to the London area, 195,000 to other cities and market towns and 1,000,000 to villages and hamlets. He used a series of multipliers, 5·4 for a house within the walls of London, 4·6 for a house within the liberties, 4·4 for the out parishes in Surrey and Middlesex and 4·3 for Westminster. For other towns, his multiplier was 4·3 and for villages 4·0.

Having performed his multiplications he gives London a bonus of 10%, other towns 2% and villages 1%. Lastly he estimates homeless people to number 80,000. The final result to the nearest round number is 5½ millions.

How King obtained his multiplier is not clear. In addition to the Hearth Office data he says he used 'the assessments on marriages, births, and burials, parish registers and other public accounts' and that from these he deduced the multipliers, but this is rather vague. He also classified the population by sex, civil state and age (under 1, under 5, under 10, under 16, above 16, above 21, above 25, above 60). How he reached these figures is not explained.

But nothing succeeds like success. As we shall see, his estimate of the total

* Prof. E. C. K. Gonner (*J.R. Statist. Soc.* 76, (1912–13), 261–97), in an interesting paper which I have largely used in writing this chapter, remarks that the 'houses' of the Hearth Office must have been really families or separate occupations as King indeed realized, and thinks that King fell into some confusion in attempting to replace families by houses. Gonner argued that the best way was to proceed on the basis that the Hearth Office unit of a family should be retained and be corrected for empty houses, blacksmiths' shops, etc. on the basis of 1801 census returns and the multiplier used should be persons per family of 1801. The result is to give a figure about a quarter of a million larger than King's. The method described in the text also leads to the conclusion that King somewhat understated the population.

population is probably very near the truth and Prof. Westergaard has remarked that, judging from Swedish observations of a few years later, King's age distribution is quite reasonable.

As a statistical prophet King was no more successful than his contemporaries (and successors). He believed that down to his time the population of England had doubled in 435 years, that the next doubling would require from 1200 to 1300 years and that in A.D. 3500–3600 the population would reach 22 millions of souls, in case, as he cautiously adds 'the world should last so long'. His estimates as a matter of arithmetical curiosity are excellently fitted by a logistic with an upper asymptote of fifteen millions and would give the present population as about eight millions.

Modern statisticians, such as Farr and Brownlee, have confirmed King's estimate of the population at the end of the seventeenth century in the following way. After 1801 the population was known by actual counting and for the first forty years of the nineteenth century baptisms and burials were still the only data of births and deaths. If one started from, say, the enumeration of 1831 and worked back to the population of 1821 by adding the numbers of burials and subtracting the number of baptisms then, if these really measured deaths and births, the result ought to agree with the census enumeration, provided immigration and emigration balanced. But the burials and baptisms understated deaths and births. One might adjust the figures by multipliers to bring the result into agreement with the census and then test against another backward run of ten years. Brownlee found that if the number of burials were multiplied by 1·2 and the number of baptisms by 1·243, the agreement was good.

This may seem a highly conjectural method, but it certainly gives quite good results. The difference between births and deaths estimated in this way for the decennium 1801–10, I find to be about 12·4 per 1000 living. If one multiplies the enumerated population of 1801 by $(1{\cdot}0124)^{10}$ we reach 10·1 millions, not a bad approximation to 10·2 millions actually counted. Assuming that before 1801 burials and baptisms had the same relation to deaths and births as between 1801 and 1841, we can work backwards to the beginning of the eighteenth century with the result that the population then was about 5·8 millions, not much more than King's estimate. In view of the following discussion it will be useful to consider the probable state of the population (as determined by these methods) in the eighteenth century. In the first sixty years of the century it grew very slowly, was about 6·1 millions in 1751 and 6·5 millions in 1761. It then began to increase faster, was 7·5 in 1781, 8·2 by 1791 and 9·2 at the census of 1801 (8·9 as enumerated, but an estimate of a deficit of 1/30th was made).

From Gregory King's time to the census of 1801 we have a series of more or less intelligent guesses.

These are well described in Prof. Gonner's paper.* Two schools of thought

* *J.R. Statist. Soc.* 76 (1912–13), 261–96.

did battle in the eighteenth century; the pessimists who held that the population was decreasing and the country going steadily to the dogs, and the optimists who believed just the contrary. Both used the same weapons. The heavy artillery was a return of houses for taxation purposes increased conjecturally by a figure for houses which escaped taxation, the sum multiplied by a conjectural average of persons per house. As light artillery one had the yield of taxes on commodities and the returns of baptisms and burials.

A pessimist put the number of untaxed houses low and the multiplier low, and an optimist raised both.

The first controversy which took place in 1754 in the proceedings of the Royal Society did not attract much notice. Brackenridge (mildly pessimistic) pointed out that the number of houses assessed to house tax had decreased between 1710 and 1754 from 729,048 to 690,000, which suggested a decrease of population (by a previous conjectural calculation based on burials and baptisms, he had reckoned a small increase, which was probably correct). Much turned on the number of houses which did not pay tax (either because the occupant was in receipt of alms, did not, owing to poverty, contribute to the church or poor rate, or through mere default). Brackenridge put the number at 200,000. His critic, Forster, argued that Brackenridge under-stated the number of untaxed houses, adducing a sample of nine country parishes with 588 houses of which only 177 were taxed and a market town with 229 taxed houses out of 448. Using these figures as a basis for conjecture Forster raises Brackenridge's 890,000 to 1,427,110. From this (with a multiplier of 6 for town houses and 5 for country houses) he reaches a population of seven and a half millions—probably a considerable over-estimate.

The next controversy was a quarter of a century later (in a period when the population was certainly increasing) and its originator was Dr Richard Price (1723–91), who has attained a posthumous celebrity reminiscent of the man whose title to distinction was that he had once been kicked by George IV. Most readers know him as the preacher of a sermon which was the text of Burke's *Reflections*, most students of economic history know him as the inventor of that theory of the virtue of a Sinking Fund which has been likened to the economic system of a community which prospered by taking in one another's washing; most vital statisticians remember him as the computor of the Northampton Life Table which gave a seriously incorrect picture of prevailing mortality and indirectly cost the country a large sum of money. Finally, in the controversy about to be described, Price was pertinaciously in the wrong on all the main issues.

The apparent inference from all this is that Price was either a fool or a knave. Gainsborough's portrait of the Rev. Richard Price, which hangs (or did hang) in the Board Room of the Equitable Assurance Society, gives no support to the hypothesis that Price was a fool; his life would be a promising field of research for a young historian with a competent knowledge of economics. His importance

in statistical history is not great enough to justify me in a critical study (even if I had the necessary training in finance and economics). My guess is that Price was an able, self-confident, original-minded man, who knew a good deal about many things and had no exact knowledge of anything. He had 'a way' with him, he could *interest* people. In fact he had some of the qualities of Petty. It is easy enough to make jokes about his notion of the mysterious power of money to increase at compound interest and it is possible that William Pitt the younger (who was only a boy when he adopted Price's theory) was not a good economic reasoner. Still, even 150 years ago, there were bankers and Treasury officials, and it is possible that both they (and Price) were not so much bad theoretical reasoners as shrewd opportunists, that they were deliberately blind to the speciousness of an attractive defence of a desirable financial expedient. I have myself sometimes wondered whether, in the eighteenth century, an Assurance Society would have minded very much if a Life Table had erred on the pessimistic side.

Price did not enter on the population question with an unbiased mind. He was a keen politician and he believed that the policy of the government was bad for the country; he also believed that the wealth of a country was its people. Hence he believed that the population was declining and nothing shook that belief. Had he survived another ten years, until the first census, he would probably have disputed the accuracy of the returns.

Price began with the figures of houses in 1690, which he cited from Davenant (they were really due to King, who communicated them to Davenant), making the total 1,319,000. He then gave the figures of assessed, chargeable and cottages (cottages being houses too small to be taxed) as 678,915, 25,628 and 276,149, making a total of 980,692 in 1761. In 1777 they were 682,077, 19,396 and 251,261, a total of 952,734. On this basis he concluded that the population had declined by about one and a half millions and was actually less than five millions.

Howlett and Wales, Price's chief opponents, impugned every step in the reasoning. First, they pointed out that in the estimate for 1690 there was almost certainly a confusion between families and houses. Then they argued that many householders evaded duty (for instance by the simple plan of blocking up windows (the prayer 'Lighten our darkness, we beseech thee, Oh Pitt' is still remembered) and showed by direct enumeration in certain parishes that the returns were inaccurate. Finally, they gave reason to think that Price's multiplier was too small. On each of these points they were probably right. Indeed Price was obliged to admit the validity of some of their criticisms. But he declined to budge; sometimes he took *ad captandum* advantage of arithmetical slips by his adversaries, sometimes he declined to admit that their samples were representative, sometimes he tried to ignore the effect of corrections which he was forced to make.

These were the principal arguments. Both parties used the data of burials

and baptisms as subsidiary arguments. Price seems only to have used the London Bills, which rather let him down; because although they seemed to help for some part of the century, he admits that by 1773 London was increasing and, very characteristically, uses this as in his favour: 'But it appears that, in truth, this is an event more to be dreaded than desired. The more London increases, the more the rest of the country must be deserted.' Price's adversaries went farther afield and counted burials and christenings in 162 parishes in all parts of England for two quinquennia, one beginning in 1758, the other in 1773. Baptisms increased from 47,638 to 59,567, burials from 49,553 to 53,030.

But neither party put much weight upon what we should now consider primary evidence; rightly, because of its incompleteness.

But these data were not wholly neglected by medical writers as we shall see in later sections. One may fairly say on the evidence here summarized that the eighteenth-century political arithmeticians of England made no advance whatever upon the position reached by Graunt, Petty and King. They were second-rate imitators of men of genius.

(To be concluded)

MOMENTS OF THE DISTRIBUTIONS OF POWERS AND PRODUCTS OF NORMAL VARIATES

By J. B. S. HALDANE, F.R.S.

CONTENTS

1. Introductory

When numerous organisms or organs are weighed, the distribution of the weights is often positively skew. On the other hand, the distributions of linear measurements is often very close to normal. The question then arises, given that the volume of an organ is proportional to the product of three mutually perpendicular measurements, each being normally distributed, what will be the distribution of the volumes? In general the three linear measurements will be correlated, and the problem might appear hopelessly complex. However, it will be shown later that, provided certain conditions are fulfilled, the distributions all lie very close together.

The problem can obviously be generalized to cover the case of the product of any number of normal variates. The most interesting cases are those of two, three, or an infinite number of such variates. Further, two special cases are comparatively simple. When the coefficients of correlation are all equal to unity and the coefficients of variation equal we are concerned with the distribution of a power of the normal variate. When the correlations all vanish, we are concerned with that of a product of several uncorrelated normal variates.

It is not, of course, suggested that all skew variation of weights is to be explained on these lines. For example the Galton-Macalister distribution, in which the logarithm of the variate is normally distributed, can be thought of as arising in at least two different ways. The weight may be the product of a large number of normal variates; or for constant cell size, the number of cell generations

may be distributed in a certain manner about a mean in each organ, these numbers being normally distributed. The highly skew variation of human weights is probably to be explained by the fact that a rather small fraction of the human race lays down very large quantities of fat. Nevertheless, it will be shown that simple criteria will determine whether observed positive skewness and leptokurtosis, too large to be ascribed to sampling error, can be explained on the lines discussed above. And in any particular case it is worth while finding out whether this is so.

Any measure of asymmetry, such as $\gamma_1 = \sqrt{\beta_1}$, or of kurtosis, such as $\gamma_2 = \beta_2 - 3$, is a dimensionless number independent of the unit of measurement. Hence in a transformed normal distribution of the type here considered it must clearly be a function of the only dimensionless number derivable from the first two moments, namely, the coefficient of variation, c. In what follows we shall generally use m for the mean of the original normal distribution and km^2 for its variance, so that $k^{\frac{1}{2}}$ is the coefficient of variation. The usual notation is used for the mean, variance and other moments and cumulants of the derived distribution.

The distribution of the product of a pair of correlated normal variates has already been fully discussed by Craig (1936). If m_1 and m_2 are the mean values of X and Y, k_1 and k_2 their coefficients of variation, and ρ their coefficients of correlation, then Craig finds for the cumulant generating function of

$$\frac{XY}{m_1 m_2 \sqrt{(k_1 k_2)}},$$

$$\kappa(\theta) = \frac{\theta\left[\left(\frac{1}{k_1}+\frac{1}{k_2}-\frac{2\rho}{\sqrt{(k_1k_2)}}\right)\theta+\frac{2}{\sqrt{(k_1k_2)}}\right]}{2[1-(1+\rho)\theta][1+(1-\rho)\theta]} - \tfrac{1}{2}\log[1-(1+\rho)\theta] - \tfrac{1}{2}\log[1+(1-\rho)\theta].$$

If $k_1 = k_2 = k$, this becomes

$$\kappa(\theta) = \frac{\theta}{k[1-(1+\rho)\theta]} - \tfrac{1}{2}\log[1-(1+\rho)\theta] - \tfrac{1}{2}\log[1+(1-\rho)\theta].$$

The rth cumulant of $\dfrac{XY}{m_1 m_2\sqrt{(k_1k_2)}}$ is

$$\kappa_r = \tfrac{1}{4}r!\left[\{(1+\rho)^{n-1}-(\rho-1)^{n-1}\}\left(\frac{1}{k_1}+\frac{1}{k_2}-\frac{2\rho}{\sqrt{(k_1k_2)}}\right)\right.$$
$$\left.+\{(1+\rho)^n-(\rho-1)^n\}\frac{2}{\sqrt{(k_1k_2)}}\right]+\tfrac{1}{2}(r-1)!\,[(1+\rho)^n+(\rho-1)^n].$$

If $k_1 = k_2 = k$, the rth cumulant of XY is

$$\kappa_r = (m_1m_2)^r k^{r-1}[(1+\rho)^{r-1}r! + \tfrac{1}{2}k\{(1+\rho)^r+(\rho-1)^r\}(r-1)!].$$

Craig's discussion is mainly confined to the cases where k_1 and k_2 are large. In the cases of greatest biometrical interest they are small, and some points of interest arise, besides those dealt with by Craig.

2. Distribution of the cube of a normal variate

If x be a reduced normal variate, that is to say, a variate whose mean is zero and standard deviation unity, it is required to find the distribution of X^3, where $X = m(1+k^{\frac{1}{2}}x)$. We note that

$$\overline{x^{2r+1}} = 0, \text{ and } \overline{x^{2r}} = (2r-1)(2r-3)(2r-5)\ldots 3.1 = \frac{(2r)!}{2^r.r!}.$$

Hence

$$\overline{X^3} = m^3(1+3k^{\frac{1}{2}}\overline{x}+3k\overline{x^2}+k^{\frac{3}{2}}\overline{x^3}) = m^3(1+3k),$$

$$\overline{X^6} = m^6(1+6k^{\frac{1}{2}}\overline{x}+15k\overline{x^2}+20k^{\frac{3}{2}}\overline{x^3}+15k^2\overline{x^4}+6k^{\frac{5}{2}}\overline{x^5}+k^3\overline{x^6})$$
$$= m^6(1+15k+45k^2+15k^3),$$

so that the first four moments of X^3 about zero are

$$\mu_1' = m^3(1+3k),$$
$$\mu_2' = m^6(1+15k+45k^2+15k^3),$$
$$\mu_3' = m^9(1+36k+378k^2+1260k^3+945k^4),$$
$$\mu_4' = m^{12}(1+66k+1485k^2+13860k^3+51975k^4+62370k^5+10395k^6).$$

Hence the moments about the mean $m^3(1+3k)$, and the cumulants, are

$$\left.\begin{aligned}
\kappa_1 = \mu_1' &= m^3(1+3k),\\
\kappa_2 = \mu_2 &= 3m^6k(3+12k+5k^2),\\
\kappa_3 = \mu_3 &= 54m^9k^2(3+16k+15k^2),\\
\mu_4 &= 27m^{12}k^2(9+240k+1326k^2+1920k^3+385k^4),\\
\kappa_4 \quad &= 648m^{12}k^3(7+48k+75k^2+15k^3).
\end{aligned}\right\} \tag{1}$$

It is to be noted that in these calculations, and in the majority, though not all, those of this paper, the expressions for the moments about the mean are simpler than those for the moments about zero. Successive moments about the mean are therefore best calculated from the former, so far as possible, rather than the latter. Thus the equation

$$\mu_4 = \mu_4' - \mu_1'(4\mu_3+6\mu_1'\mu_2+\mu_1'^3),$$

involves less algebra than the more usual

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_1'^2\mu_2 - 3\mu_1'^4.$$

It follows that

$$\sigma = 3m^3k^{\frac{1}{2}}(1+4k+\tfrac{5}{3}k^2)^{\frac{1}{2}},$$

$$c = \frac{(9k+36k^2+15k^3)^{\frac{1}{2}}}{1+3k},$$

$$\gamma_1 = \frac{+6\sqrt{(3k)}\,(3+16k+15k^2)}{(3+12k+5k^2)^{\frac{3}{2}}},$$

$$\gamma_2 = \frac{+72k(7+48k+75k^2+15k^3)}{(3+12k+5k^2)^2}.$$

Hence the distribution is positively skew, and leptokurtic. If c is small, we have, approximately,

$$\left.\begin{aligned}\gamma_1 &= 2c+\frac{2c^3}{27}+\ldots,\\ \gamma_2 &= \frac{56c^2}{9}+\frac{16c^4}{27}+\ldots.\end{aligned}\right\} \tag{2}$$

Since in most practical cases $c^2 < 0{\cdot}01$, the first term will have an error of less than $0{\cdot}1\,\%$, and no attention need be paid to the later terms.

3. Distribution of any power of a normal variate

It is required to find the moments of $m^n(1+k^{\frac{1}{2}}x)^n$. The mean value of its rth power is clearly

$$\mu'_r = m^{nr}\left[1+\frac{(nr)!\,k}{2(nr-2)!}+\frac{(nr)!\,k^2}{8(nr-4)!}+\ldots-\frac{(nr)!\,k^s}{2^s\,.\,s!(nr-2s)!}+\ldots\right].$$

This converges for all positive values of n, since k is small, though it only terminates for positive integral values of nr. After somewhat tedious algebra, we find for the mean and other moments and cumulants:

$$\left.\begin{aligned}\kappa_1 = \mu'_1 &= m^n[1+\tfrac{1}{2}n(n-1)k+\ldots],\\ \kappa_2 = \mu_2 &= m^{2n}n^2k[1+\tfrac{1}{2}(n-1)(3n-5)k+\ldots],\\ \kappa_3 = \mu_3 &= 3m^{3n}n^3(n-1)k^2[1+\tfrac{1}{8}(17n^2-55n+44)k+\ldots],\\ \mu_4 &= 3m^{4n}n^4k^2[1+\tfrac{5}{3}(n-1)(5n-7)k+\ldots],\\ \kappa_4 \quad &= 4m^{4n}n^4(n-1)(4n-5)k^3[1+0(k)].\end{aligned}\right\} \tag{3}$$

Hence
$$\begin{aligned}c &= nk^{\frac{1}{2}}[1+\tfrac{1}{4}(n-1)(3n-5)k+\ldots],\\ \gamma_1 &= 3(n-1)k^{\frac{1}{2}}[1+\tfrac{1}{12}(7n^2-38n+43)k+\ldots],\\ \gamma_2 &= 4(n-1)(4n-5)k[1+0(k)].\end{aligned}$$

And when c is small

$$\left.\begin{aligned}\gamma_1 &= \frac{3(n-1)}{n}c[1+0(c^2)],\\ \gamma_2 &= \frac{4(n-1)(4n-5)c^2}{n^2}[1+0(c^2)].\end{aligned}\right\} \tag{4}$$

When $n = 2$, the expressions become very simple,

$$\left.\begin{aligned}\kappa_1 &= m^2(1+k),\\ \kappa_2 &= 2m^4k(2+k),\\ \kappa_3 &= 8m^6k^2(3+k),\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ \kappa_r &= (r-1)!\,m^{2r}(2k)^{r-1}(r+k).\end{aligned}\right\} \tag{5}$$

This is in accordance with Craig's formula, putting $k_1 = k_2 = k$, $\rho = 1$.

4. Distribution of the product of three independent normal variates

Let x, y, z, be independent reduced normal variates, so that $\overline{x^2} = \overline{y^2} = \overline{z^2} = 1$, $\overline{x^4} = \overline{y^4} = \overline{z^4} = 3$, etc. Let $X = m_1(1+k^{\frac{1}{2}}x)$, $Y = m_2(1+k^{\frac{1}{2}}y)$, $Z = m_3(1+k_3^{\frac{1}{2}}z)$. Here $k_1^{\frac{1}{2}}$, $k_2^{\frac{1}{2}}$, $k_3^{\frac{1}{2}}$ are coefficients of variation of linear measurements, so that k_1, k_2, k_3, are commonly about 0·001. It is required to find the moments of the distribution of XYZ. In the expansion of a power of XYZ we need only consider even powers of x, y and z. For example,

$$\begin{aligned}\overline{X^2Y^2Z^2} &= m_1^2m_2^2m_3^2\overline{(1+k_1^{\frac{1}{2}}x)^2(1+k_2^{\frac{1}{2}}y)^2(1+k_3^{\frac{1}{2}}z)^2}\\ &= m_1^2m_2^2m_3^2(1+k_1\overline{x^2})(1+k_2\overline{y^2})(1+k_3\overline{z^2})\\ &= m_1^2m_2^2m_3^2(1+k_1)(1+k_2)(1+k_3).\end{aligned}$$

So if $m_1m_2m_3 = V$ (a volume), the moments of XYZ about zero are

$$\begin{aligned}\mu_1' &= V,\\ \mu_2' &= V^2(1+k_1)(1+k_2)(1+k_3),\\ \mu_3' &= V^3(1+3k_1)(1+3k_2)(1+3k_3),\\ \mu_4' &= V^4(1+6k_1+3k_1^2)(1+6k_2+3k_2^2)(1+6k_3+3k_3^2).\end{aligned}$$

So

$$\begin{aligned}\kappa_2 = \mu_2 &= V^2(\Sigma k_1+\Sigma k_1k_2+k_1k_2k_3),\\ \kappa_3 = \mu_3 &= 6V^3(\Sigma k_1k_2+4k_1k_2k_3),\\ \mu_4 &= 3V^4(\Sigma k_1^2+2\Sigma k_1k_2+6\Sigma k_1^2k_2+38k_1k_2k_3+3\Sigma k_1^2k_2^2\\ &\qquad +36\Sigma k_1^2k_2k_3+18\Sigma k_1^2k_2k_3+9k_1^2k_2^2k_3^2,\\ \kappa_4 &= 6V^4[2\Sigma k_1^2k_2+\Sigma k_1^2k_2^2+4k_1k_2k_3(4+4\Sigma k_1+2\Sigma k_1k_2+k_1k_2k_3)].\end{aligned}$$

In practice we can neglect all terms except the leading one, and write

$$\left.\begin{aligned}\kappa_1 &= V,\\ \kappa_2 &= V^2(k_1+k_2+k_3),\\ \kappa_3 &= 6V^3(k_2k_3+k_3k_1+k_1k_2),\\ \kappa_4 &= 12V^4(k_1^2k_2+k_2^2k_3+k_3^2k_1+k_1k_2^2+k_2k_3^2+k_3k_1^2+8k_1k_2k_3).\end{aligned}\right\} \quad (6)$$

If $k_1 = k_2 = k_3 = k$, we find

$$\left.\begin{aligned}\kappa_1 &= V,\\ \kappa_2 &= V^2k(3+3k+k^2),\\ \kappa_3 &= 6V^3k^2(3+4k),\\ \kappa_4 &= 6V^4k^3(28+51k+24k^2+4k^3),\end{aligned}\right\} \quad (7)$$

which may be compared with equations (1). Thus

$$c = \sqrt{(3k+3k^2+k^3)},$$

$$\left.\begin{aligned}\gamma_1 &= 2c-\frac{4c^3}{9}+\dots,\\ \gamma_2 &= \frac{56c^2}{9}-\frac{22c^4}{9}+\dots\end{aligned}\right\} \quad (8)$$

It will be noticed that the leading terms of equations (2) and (8) are identical. Hence the two distributions will, in practice, be quite indistinguishable, even if tens of thousands of individuals are weighed. It follows that the cube root of the product of three independent variates is almost normally distributed. Further, the relations of γ_1 and γ_2 to c are little altered when k_1, k_2 and k_3 are different, provided that they are of the same order of magnitude. Thus if $k_1 = 0{\cdot}001$, $k_2 = 0{\cdot}002$, $k_3 = 0{\cdot}003$, we find $\gamma_1 = \frac{11}{3}c$ instead of $2c$, and $\gamma_2 = \frac{16c^2}{3} = 5{\cdot}\dot{3}c^2$ instead of $6{\cdot}\dot{2}c^2$. Only a very large sample indeed would reveal departures of this order from the simpler expressions of equations (2) and (8).

5. Distribution of the product of n independent normal variates

Let $x_1, x_2, x_3, \ldots, x_r, \ldots, x_n$, be n independent reduced normal variates. The general case is of course somewhat complicated, so we shall only investigate the special case where all the coefficients of variation are equal.

Then if $X_r = m_r(1+k^{\frac{1}{2}}x_r)$, and $M = \Pi m_r$, the moments about zero of the distribution of the product ΠX_r are

$$\begin{aligned}
\mu_1' &= M,\\
\mu_2' &= M^2(1+k)^n,\\
\mu_3' &= M^3(1+3k)^n,\\
\mu_4' &= M^4(1+6k+3k^2)^n,\\
\mu_5' &= M^5(1+10k+15k^2)^n,\\
\mu_6' &= M^6(1+15k+45k^2+15k^3)^n.
\end{aligned}$$

Hence

$$\begin{aligned}
\mu_2 &= M^2[(1+k)^n-1],\\
\mu_3 &= M^3[(1+3k)^n-3(1+k)^n+2],\\
\mu_4 &= M^4[(1+6k+3k^2)^n-4(1+3k)^n+6(1+k)^n-3],\\
\mu_5 &= M^5[(1+10k+15k^2)^n-5(1+6k+3k^2)^n+10(1+3k)^n-10(1+k)^n-4],\\
\mu_6 &= M^6[(1+15k+45k^2+15k^3)^n-6(1+10k+15k^2)^n+15(1+6k+3k^2)^n\\
&\qquad -20(1+3k)^n+15(1+k)^n-5],
\end{aligned}$$

$$\begin{aligned}
\kappa_4 &= M^4[(1+6k+3k^2)^n-3(1+2k+k^2)^n-4(1+3k)^n+12(1+k)^n-6],\\
\kappa_5 &= M^5\{(1+10k+15k^2)^n+5[-(1+6k+3k^2)^n-2(1+4k+3k^2)^n\\
&\qquad +6(1+2k+k^2)^n+4(1+3k)^n-12(1+k)^n]+24\},\\
\kappa_6 &= M^6\{(1+15k+45k^2+15k^3)^n+15[2(1+3k+3k^2+k^3)^n\\
&\qquad -(1+7k+9k^2+3k^3)^n+2(1+6k+3k^2)^n+8(1+4k+3k^2)^n\\
&\qquad -18(1+2k+k^2)^n-8(1+3k)^n+24(1+k)^n-8]\\
&\qquad -6(1+10k+15k^2)^n-10(1+6k+9k^2)^n\}.
\end{aligned}$$

The higher cumulants can be evaluated as follows:

$$\kappa_3 = 3M^3 \sum_{r=1}^{\infty} \binom{n}{r} (3^{r-1}-1)\,k^r,$$

$$\kappa_4 = 3M^4 \sum_{r=1}^{\infty} \binom{n}{r} (3^{r-1}-1)\,[(2+k)^r-4]\,k^r, \text{ etc.}$$

Thus we find

$$\left.\begin{aligned}
\kappa_1 &= M,\\
\kappa_2 &= M^2nk[1+\tfrac{1}{2}(n-1)\,k+\ldots],\\
\kappa_3 &= M^3n(n-1)\,k^2[3+4(n-2)\,k+\ldots],\\
\kappa_4 &= M^4n(n-1)\,k^3[4(4n-5)+3(13n^2-49n+47)\,k+\ldots],\\
\kappa_5 &= 5M^5n(n-1)\,k^4[(5n-6)\,(5n-7)+2(n-2)\,(143n-345)\,(k+\ldots],\\
\kappa_6 &= 3M^6n(n-1)\,k^5[432n^3-1853n^2+2917n-1758+0(k)].
\end{aligned}\right\} \quad (9)$$

These may be compared with equations (3).

$$c = \sqrt{(nk)}\,[1+\tfrac{1}{4}(n-1)\,k+\ldots],$$

and when c is small

$$\gamma_1 = \frac{3(n-1)\,c}{n}\,[1+0(c^2)],$$

$$\gamma_2 = \frac{4(n-1)\,(4n-5)\,c^2}{n^2}\,[1+0(c^2)],$$

as in equations (4).

When $n=2$ we have the simple forms

$$\begin{aligned}
\kappa_1 &= M,\\
\kappa_2 &= M^2k(2+k),\\
\kappa_3 &= 6M^3k^2,\\
\kappa_4 &= 6M^4k^3(4+k),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\kappa_{2r} &= (2r-1)!\,M^{2r}k^{2r-1}(2r+k),\\
\kappa_{2r+1} &= (2r+1)!\,M^{2r+1}k^{2r}.
\end{aligned}$$

This is in accordance with Craig's formula, putting $k_1 = k_2 = k$, $\rho = 0$.

6. Distribution of the product of two correlated normal variates

In general the different linear dimensions of an organ or organism are positively correlated. Organic correlations may reach very high values, such as 0·9, and presumably even higher values would be found for two approximately equal diameters of an approximate solid of revolution, such as an apple or an egg, where ρ may be taken as unity. On the other hand, quite low values are found. Thus the

length and breadth of a homogeneous group of like-sexed adult human skulls generally show a correlation of about 0·3 or 0·4. Hence the area of an organ will commonly be proportional to the product of two correlated variables, the volume to the product of three.

Let x and y be two reduced normal variates, with correlation ρ. The cumulant-generating function for their joint distribution is $\frac{1}{2}(t^2+2\rho tu+u^2)$. That is to say if $r+s$ is odd, then $\overline{x^r y^s}=0$. If $r+s$ is even, then $\overline{x^r y^s}$ is the coefficient of $\dfrac{t^r u^s}{r!\,s!}$ in $\exp\frac{1}{2}(t^2+2\rho tu+u^2)$. That is to say, $\overline{x^r y^s}$ is the coefficient of $t^r u^s$ in

$$(t^2+2\rho tu+u^2)^{\frac{1}{2}(r+s)},$$

multiplied by

$$\frac{r!\,s!}{2^{\frac{1}{2}(r+s)}[\frac{1}{2}(r+s)]!}.$$

It follows that

$$\begin{array}{llll}
\overline{x^2}=1, & \overline{xy}=\rho, & & \\
\overline{x^4}=3, & \overline{x^3y}=3\rho, & \overline{x^2y^2}=1+2\rho^2, & \\
\overline{x^6}=15, & \overline{x^5y}=15\rho, & \overline{x^4y^2}=3(1+4\rho^2), & \overline{x^3y^3}=3\rho(3+2\rho^2),\\
\overline{x^8}=105, & \overline{x^7y}=105\rho, & \overline{x^6y^2}=15(1+6\rho^2), & \overline{x^5y^3}=15\rho(3+4\rho^2),\\
 & & \overline{x^4y^4}=3(3+24\rho^2+8\rho^4), &
\end{array}$$

$$\overline{x^8y^2}=105(1+8\rho^2),\qquad \overline{x^7y^3}=315\rho(1+2\rho^2),\qquad \overline{x^6y^4}=45(1+12\rho^2+8\rho^4),$$

$$\overline{x^8y^4}=315(1+16\rho^2+16\rho^4).$$

Of course $\overline{x^s y^r}=\overline{x^r y^s}$.

Thus the moments of xy about its mean ρ are

$$\mu_2=1+\rho^2,\quad \mu_3=2\rho(3+\rho^2),\quad \mu_4=3(3+14\rho^2+3\rho^4).$$

Hence $\kappa_4=6(1+6\rho^2+\rho^4)$. Hence the distribution is leptokurtic, and asymmetrical unless ρ vanishes.

Now consider two correlated normal variates

$$X=m_1(1+ax),\quad\text{and}\quad Y=m_2(1+by),$$

where a and b are coefficients of variation. Let $m_1m_2=A$. Then the moments of XY about zero are

$$\mu'_1=A(1+\rho ab),$$

$$\mu'_2=A^2[1+a^2+4\rho ab+b^2+(1+2\rho^2)a^2b^2],$$

$$\mu'_3=A^3[1+3(a^2+3\rho ab+b^2)+9ab\{\rho a^2+(1+2\rho^2)ab+\rho b^2\}+3\rho(3+2\rho^2)a^3b^3],$$

$$\begin{aligned}\mu'_4=A^4[&1+2(3a^2+8\rho ab+3b^2)+3\rho a^4+16\rho a^3b+12(1+2\rho^2)a^2b^2+16\rho ab^3+3\rho b^4\\&+6a^2b^2\{3(1+4\rho^2)a^2+8\rho(3+2\rho^2)ab+3(1+4\rho^2)b^2\}+3(3+24\rho^2+8\rho^4)a^4b^4].\end{aligned}$$

Hence

$$\kappa_2 = \mu_2 = A^2[a^2+2\rho ab+b^2+(1+\rho^2)a^2b^2],$$
$$\kappa_3 = \mu_3 = 2A^3ab[3\{\rho a^2+(1+\rho^2)ab+\rho b^2\}+\rho(3+\rho^2)a^2b^2],$$
$$\mu_4 = 3A^3[(a^2+2\rho ab+b^2)^2+2a^2b^2\{(3+7\rho^2)a^2+2\rho(7+3\rho^2)ab + (3+7\rho^2)b^2\}+(3+14\rho^2+3\rho^4)a^4b^4],$$
$$\kappa_4 = 6A^3a^2b^2[2\{(1+3\rho^2)a^2+2\rho(3+\rho^2)ab+(1+3\rho^2)b^2\} + (1+6\rho^2+\rho^4)a^2b^2], \quad (10)$$

in accordance with Craig's formula.

The most interesting case arises when $a = b = k^{\frac{1}{2}}$. This case is important because in practice the coefficients of variation of different linear dimensions of the same organ are often nearly equal. Thus those of linear skull measurements in like-sexed adults in a racially homogeneous population are about 0·03, so that k is about 0·001. The moments and cumulants of XY are then

$$\kappa_1 = A(1+\rho k),$$
$$\kappa_2 = \mu_2 = A^2k[2(1+\rho)+(1+\rho^2)k],$$
$$\kappa_3 = \mu_3 = 2A^3k^2[3(1+\rho)^2+(3+\rho^2)k];$$
$$\mu_4 = 3A^4k^2[4(1+\rho)^2+4(3+7\rho+7\rho^2+3\rho^3)k+(3+14\rho^2+3\rho^4)k^2],$$
$$\kappa_4 = 6A^4k^3[4(1+\rho)^3+(1+6\rho^2+\rho^4)k]. \quad (11)$$

Hence
$$c = \sqrt{[2(1+\rho)k]}\left[1+\frac{1-4\rho-3\rho^2}{4(1+\rho)}k+\ldots,\right]$$
while
$$\gamma_1 = \tfrac{3}{2}c[1+0(c^2)],$$
$$\gamma_2 = 3c^2[1+0(c^2)],$$

exactly as found when the two variates are in a constant ratio or quite independent. When however $a \neq b$, this is no longer the case, for it is clear that the distribution becomes normal when a or b vanishes. If $\frac{a^2+b^2}{2ab} = p$, so that $p \geqslant 1$, we have, for equations (10),

$$\gamma_1 = \frac{1+2p\rho+\rho^2}{(p+\rho)^2}\frac{3c}{2}, \quad \gamma_2 = \frac{1+3p\rho-3\rho^2+p\rho^3}{(p+\rho)^3}3c^2,$$

both approximately. However, a and b may differ considerably without any very great effect of γ_1/c or γ_2/c. Thus if $a = 2b$, so that $p = \frac{5}{4}$,

$$\gamma_1 = \frac{8(2+5\rho+2\rho^2)}{(5+4\rho)^2}\frac{3c}{2}.$$

That is to say, γ_1 is 64 % of $3c/2$ if $\rho = 0$, and 86 % if $\rho = \frac{1}{2}$, whilst γ_2 is 51 % of $3c$ if $\rho = 0$, and 71 % if $\rho = \frac{1}{2}$.

So far we have assumed that ρ is not negative. If $\rho = -1$, $XY = 1-kx^2$, so the mean is $A(1-k)$, and the other cumulants are given by

$$\kappa_r = (-)^r\,2^{r-1}(r-1)!\,A^rk^r.$$

If $\rho = -1 + 2\sqrt{\frac{k}{3}} + \frac{k}{2} + \dots$, κ_3 in equations (11) vanishes, though the distribution does not become quite symmetrical. However, negative values of ρ are of no biological interest.

7. DISTRIBUTION OF THE PRODUCT OF THREE CORRELATED NORMAL VARIATES

Let x, y, z, be three reduced normal variates as before, their correlations being $\rho_{yz} = \kappa$, $\rho_{zx} = \lambda$, $\rho_{xy} = \mu$. Thus the cumulant-generating function is

$$\tfrac{1}{2}(t^2 + u^2 + v^2 + 2\kappa uv + 2\lambda vt + 2\mu tu).$$

Odd moments vanish, and the even moment $\overline{x^p y^q z^r}$ is the coefficient of $t^p u^q v^r$ in

$$(t^2 + u^2 + v^2 + 2\kappa uv + 2\lambda vt + 2\mu tu)^{\frac{1}{2}(p+q+r)},$$

multiplied by
$$\frac{p!\,q!\,r!}{2^{\frac{1}{2}(p+q+r)}\,[\frac{1}{2}(p+q+r)]!}.$$

The required moments of products of two variates, such as

$$\overline{xy} = \mu, \quad \overline{x^4y^2} = 3(1+4\mu^2),$$

have already been given. The required moments of products of all three are:

$$\begin{aligned}
\overline{x^2yz} &= \kappa + 2\lambda\mu,\\
\overline{x^4yz} &= 3(\kappa + 4\lambda\mu), \quad \overline{x^3y^2z} = 3(\lambda + 2\kappa\mu + 2\kappa\mu^2), \quad \overline{x^2y^2z^2} = 1 + 2\Sigma\kappa^2 + 8\kappa\lambda\mu,\\
\overline{x^4y^3z} &= 3(3\kappa + 12\lambda\mu + 12\kappa\mu^2 + 8\lambda\mu^3),\\
\overline{x^4y^2z^2} &= 3(1 + 2\kappa^2 + 4\lambda^2 + 4\mu^2 + 16\kappa\lambda\mu + 8\lambda^2\mu^2),\\
\overline{x^3y^3z^2} &= 3(3\mu + 6\kappa\lambda + 2\mu^3 + 6\kappa^2\mu + 6\lambda^2\mu + 12\kappa\lambda\mu^2),\\
\overline{x^4y^4z^2} &= 3(3\kappa + 12\kappa^2 + 12\lambda^2 + 24\mu^2 + 96\kappa\lambda\mu + 48\kappa^2\mu^2 + 48\lambda^2\mu^2 + 8\mu^4 + 64\kappa\lambda\mu^3),\\
\overline{x^4y^3z^3} &= 9(3\kappa + 12\lambda\mu + 2\kappa^3 + 12\kappa\lambda^2 + 12\kappa\mu^2 + 24\kappa^2\lambda\mu + 8\lambda^3\mu\\
&\qquad + 8\lambda^3\mu + 8\lambda\mu^3 + 24\kappa\lambda^2\mu^2),\\
\overline{x^4y^4z^4} &= 9[3 + 24\Sigma\kappa^2 + 8\Sigma\kappa^4 + 96\Sigma\lambda^2\mu^2 + 64\kappa\lambda\mu(3 + 2\Sigma\kappa^2 + 3\kappa\lambda\mu)].
\end{aligned}$$

Other moments (except those such as $\overline{xy^2z} = \lambda + 2\kappa\mu$, which are derivable by transposition) are not needed for our purposes. It follows that xyz has a symmetrical and leptokurtic distribution, with mean zero, and

$$\begin{aligned}
\kappa_2 &= \mu_2 = 1 + 2\Sigma\kappa^2 + 8\kappa\lambda\mu,\\
\kappa_4 &= 12[2 + 17\Sigma\kappa^2 + 5\Sigma\kappa^4 + 70\Sigma\lambda^2\mu^2 + 4\kappa\lambda\mu(35 + 22\Sigma\kappa^2 + 32\kappa\lambda\mu)].
\end{aligned}$$

If $X = m_1(1+ax)$, $Y = m_2(1+by)$, $Z = m_3(1+cz)$ the expressions for the higher moments are very complicated, though it can easily be shown that the mean is $m_1m_2m_3(1 + \kappa bc + \lambda ca + \mu ab)$, and the variance

$$m_1^2m_2^2m_3^2[\Sigma a^2 + 2\Sigma\kappa bc + \Sigma(1+\kappa^2)\,b^2c^2 + 2\Sigma(2\kappa + 3\lambda\mu)\,a^2bc + (1 + 2\Sigma\kappa^2 + 8\kappa\lambda\mu)\,a^2b^2c^2].$$

We shall only give further consideration to the case where all the coefficients of variation are equal. And we shall confine ourselves to two special cases of this. In the one $\kappa = \lambda = \mu = \rho$, that is to say the variates are equally, and therefore of course positively, correlated. This is not very far from the case with the human skull. In the other, $\kappa = 1$, whilst $\lambda = \mu = \rho$. This is appropriate to an approximate solid of revolution, such as many eggs and fruits, or to a regular prism such as some sponge spicules.

In the first case we have for the moments of products of three variables:

$$\overline{x^2yz} = \rho(1+2\rho),$$

$$\overline{x^4yz} = 3\rho(1+4\rho), \quad \overline{x^3y^2z} = 3\rho(1+2\rho+2\rho^2), \quad \overline{x^2y^2z^2} = 1+6\rho^2+8\rho^3,$$

$$\overline{x^4y^3z} = 3\rho(3+12\rho+12\rho^2+8\rho^3), \quad \overline{x^4y^2z^2} = 3(1+10\rho^2+16\rho^3+8\rho^4),$$

$$\overline{x^3y^3z^3} = 3\rho(3+6\rho+14\rho^2+12\rho^3),$$

$$\overline{x^4y^4z^2} = 3(3+48\rho^2+96\rho^3+104\rho^4+64\rho^5),$$

$$\overline{x^4y^3z^3} = 9\rho(3+12\rho+26\rho^2+40\rho^3+24\rho^4),$$

$$\overline{x^4y^4z^4} = 27(1+24\rho^2+64\rho^3+104\rho^4+128\rho^5+64\rho^6).$$

So if $X = m_1(1+k^{\frac{1}{2}}x)$, $Y = m_2(1+k^{\frac{1}{2}}y)$, $Z = m_3(1+k^{\frac{1}{2}}z)$, and $m_1m_2m_3 = V$, the volume which is the product of the means, then the moments of XYZ about zero are

$$\mu_1' = V(1+3\rho k),$$

$$\mu_2' = V^2[1+3(1+4\rho)k+3(1+4\rho+10\rho^2)k+(1+6\rho^2+8\rho^3)k^2],$$

$$\mu_3' = V^3[1+9(1+3\rho)k+27(1+5\rho+8\rho^2)k^2+9(3+21\rho+54\rho^2 + 62\rho^3)k^3+27\rho(3+6\rho+14\rho^2+12\rho^3)k^4],$$

$$\mu_4' = V^4[1+6(3+8\rho)k+9(13+64\rho+88\rho^2)k^2+36(9+64\rho+160\rho^2 + 152\rho^3)k^3+27(13+128\rho+448\rho^2+768\rho^3+568\rho^4)k^4 + 54(3+24\rho+144\rho^2+304\rho^3+424\rho^4+256\rho^5)k^5 + 27(1+24\rho^2+64\rho^3+104\rho^4+128\rho^5+64\rho^6)k^6].$$

Hence

$$\kappa_1 = \mu_1' = V(1+3\rho k),$$

$$\kappa_2 = \mu_2 = V^2k[3(1+2\rho)+3(1+4\rho+7\rho^2)k+(1+6\rho^2+8\rho^3)k^2],$$

$$\kappa_3 = \mu_3 = 6V^3k^2[3(1+2\rho)^2+(41+27\rho+60\rho^2+53\rho^3)k + 3\rho(4+9\rho+8\rho^2+14\rho^3)k^2],$$

$$\mu_4 = 3V^4k^2[9(1+2\rho)^2+2(37+222\rho+471\rho^2+350\rho^3)k+3(39+316\rho + 1038\rho^2+1584\rho^3+1001\rho^4)k^2+8(3+24\rho+127\rho^2+268\rho^3 + 346\rho^4+192\rho^5)k^3+9(1+24\rho^2+64\rho^3+104\rho^4+128\rho^5+64\rho^6)k^4],$$

$$\kappa_4 = 6V^4k^3[28(1+2\rho)^3+3(17+144\rho+468\rho^2+688\rho^3+411\rho^4)k + 12(2+17\rho+92\rho^2+193\rho^3+241\rho^4+130\rho^5)k^2 + 2(2+51\rho^2+140\rho^3+225\rho^4+264\rho^5+128\rho^6)k^3]. \quad (12)$$

These expressions reduce to equations (1) if $\rho = 1$, and (7) if $\rho = 0$. When k is small, $c^2 = 3(1+2\rho)\,k$,
$$\gamma_1 = 2c[1+0(c^2)],$$
$$\gamma_2 = \frac{56c^2}{9}[1+0(c^2)].$$

Thus the relation of γ_1 and γ_2 to c is almost independent of the coefficient of correlation.

We next consider the case when $z = y$, so that $\kappa = 1$, whilst $\lambda = \mu = \rho$. This is most simply solved by finding the moments of $(1+k^{\frac{1}{2}}x)(1+k^{\frac{1}{2}}y)^2$. Thus if $X = m_1(1+k^{\frac{1}{2}}x)$, and $Y = m_2(1+k^{\frac{1}{2}}y)$, the moments of XY^2 about zero are

$$\mu_1' = V[1+(1+2\rho)\,k],$$
$$\mu_2' = V^2[1+(7+8\rho)\,k+3(3+8\rho+4\rho^2)\,k^2+3(1+4\rho^2)\,k^3],$$
$$\mu_3' = V^3[1+18(1+\rho)\,k+18(5+11\rho+5\rho^2)\,k^2+30(5+15\rho+18\rho^2+4\rho^3)\,k^3 \\ +45(1+6\rho+6\rho^2+8\rho^3)\,k^4],$$
$$\mu_4' = V^4[1+2(17+16\rho)\,k+3(127+256\rho+112\rho^2)\,k^2+84(21+64\rho+64\rho^2 \\ +16\rho^3)\,k^3+105(31+128\rho+192\rho^2+128\rho^3+16\rho^4)\,k^4 \\ +630(3+16\rho+32\rho^2+32\rho^3+16\rho^4)\,k^5+315(1+16\rho^2+16\rho^4)\,k^6].$$

Hence

$$\kappa_1 = \mu_1' = V[1+(1+2\rho)\,k],$$
$$\kappa_2 = \mu_2 = V^2k[5+4\rho+4(2+5\rho+2\rho^2)\,k+3(1+4\rho^2)\,k^2],$$
$$\kappa_3 = \mu_3 = 2V^3k^2[3(8+14\rho+5\rho^2)+2(29+84\rho+87\rho^2+16\rho^3)\,k \\ +9(2+14\rho+13\rho^2+16\rho^3)\,k^2],$$
$$\mu_4 = 3V^4k^2[(5+4\rho)^2+8(40+113\rho+95\rho^2+22\rho^3)\,k+2(427 \\ +1628\rho+2376\rho^2+1376\rho^2+160\rho^3+160\rho^4)\,k^2 \\ +24(24+121\rho+237\rho^2+234\rho^3+104\rho^4)\,k^3+105(1+16\rho^2+16\rho^4)\,k^4],$$
$$\kappa_4 = 24V^4k^3[30+80\rho+65\rho^2+14\rho^3+(95+364\rho+513\rho^2+292\rho^3 \\ +32\rho^4)\,k+3(22+116\rho+227\rho^2+214\rho^3+96\rho^4)\,k^2 \\ +3(4+67\rho^2+64\rho^4)\,k^3]. \qquad (13)$$

Hence, approximately,

$$c^2 = (5+4\rho)\,k,$$
$$\gamma_1 = \frac{6(2+\rho)(4+5\rho)\,c}{(5+4\rho)^2},$$
$$\gamma_2 = \frac{24(30+80\rho+65\rho^2+14\rho^3)\,c^2}{(5+4\rho)^3}.$$

Hence γ_1 varies between $\frac{48}{25}c$, or $1{\cdot}92c$ when $\rho = 0$, and $2c$ when $\rho = 1$. This is its maximum, so for high values of ρ, γ_1/c is nearly constant. It vanishes when $\rho = -0{\cdot}8$. γ_2 increases from $\frac{144}{25}c^2$, or $5{\cdot}76c^2$, when $\rho = 0$, to $\frac{56}{9}c^2$, or $6{\cdot}\dot{2}c^2$ when

$\rho = 1$. This again is its maximum, so γ_2/c^2 is nearly constant for large ρ. γ_2 does not vanish for any admissible values of ρ. In fact for positive values of ρ it would be impossible except with enormous samples, to distinguish this distribution from that of the cube of a normal variate. If $\rho = 1$, equations (13) reduce to equations (1). If $\rho = 0$, they give the cumulants of XY^2, where X and Y are uncorrelated normal variates with the same coefficient of variation, namely,

$$\kappa_1 = V(1+k), \qquad \kappa_2 = V^2k(5+8k+3k^2),$$
$$\kappa_3 = 4V^2k^2(12+29k+9k^2), \quad \kappa_4 = 24V^4k^3(30+95k+66k^2+12k^3),$$

which may easily be obtained independently.

8. The product of n correlated normal variates

If the variates are $X_1, X_2, \ldots, X_r, \ldots, X_n$, where $X_r = m_r(1+a_rx_r)$, and $P = \Pi m_r$, while x_r is a reduced normal variate, and ρ_{rs} is the coefficient of correlation of x_r and x_s, and hence of X_r and X_s, the general expression for the moments of $X_1, X_2, \ldots, X_n$ is complicated. It can however easily be seen that the mean is:

$$P[1+\Sigma\rho_{rs}a_ra_s+\Sigma(\rho_{rs}\rho_{tu}+\rho_{rt}\rho_{su}+\rho_{ru}\rho_{ts})\,a_ra_sa_ta_u+\ldots],$$

while
$$\mu_2 = P^2[\Sigma a_r^2+2\Sigma\rho_{rs}a_ra_s+\ldots].$$

If every $a_r = k^{\frac{1}{2}}$, and every $\rho_{rs} = \rho$, then $\overline{x_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3}\ldots}$ vanishes when $m = \Sigma\alpha_r$ is odd, and when m is even it is the coefficient of $t_1^{\alpha_1}t_1^{\alpha_2}\ldots$ in the expansion of $(\Sigma t_r^2+2\rho\Sigma t_rt_s)^m$, multiplied by $\alpha_1!\,\alpha_2!\,\alpha_3!/(2^m m!)$. Thus the moments about zero are:

$$\mu_1' = P\left[1+\tfrac{1}{2}n(n-1)\,\rho k+\tfrac{1}{8}n(n-1)\,(n-2)\,(n-3)\,\rho^2k^2+\ldots+\frac{n!\,\rho^rk^r}{2^rr!\,(n-2r)!}+\ldots\right],$$
$$\mu_2' = P^2[1+n\{1+2(n-1)\,\rho\}\,k+\tfrac{1}{2}n(n-1)$$
$$\times\{1+4(n-2)\,\rho+2(2n^2-6n+5)\,\rho^2\}\,k^2+\ldots],$$
$$\mu_3' = P^3[1+3n\{1+\tfrac{3}{2}(n-1\rho\}\,k+\tfrac{9}{2}n(n-1)$$
$$\times\{1+(3n-4)\,\rho+\tfrac{1}{4}(9n^2-21n+14)\,\rho^2\}\,k^2+\ldots],$$

So

$$\mu_2 = P^2nk[1+(n-1)\,\rho+\tfrac{1}{2}(n-1)\,\{1+4(n-2)\,\rho+(3n^2-9n+7)\,\rho^2\}\,k+\ldots],$$
$$\mu_3 = 3P^2n(n-1)\,[1+(n-1)\,\rho^2k^2[1+0(k)]. \tag{14}$$

Thus
$$c = [n\{1+(n-1)\,\rho\}\,k]^{\frac{1}{2}}\,[1+0(k)],$$
$$\gamma_1 = 3(n-1)\left[\frac{1+(n-1)\,\rho}{n}\right]^{\frac{1}{2}}[1+0(k)] \tag{15}$$
$$= \frac{3(n-1)}{n}c^2[1+0(c^2)],$$

and presumably γ_2 etc. are approximately the same functions of c as in the case of equations (9).

9. The Galton-MacAlister distribution

Consider the distribution of the nth power of a normal variate when n is very large, and k very small, but $nk^{\frac{1}{2}}$ remains constant. All but a few terms of equations (3) will vanish, and if $nk^{\frac{1}{2}}$ is small, the distribution becomes identical with the distribution described by MacAlister (1879) whose first four moments have been given by Pearson (1905). Its cumulants can readily be found as follows: Given that $x = \log X$ is normally distributed, to find the distribution of X. Let m and s be the mean and standard deviation of x. Then $X = e^x$. Hence, since the moment-generating function of x is $M(t) = e^{mt+\frac{1}{2}st^2}$, the rth moment of X about zero is

$$\mu_r' = \overline{e^{rx}} = 1 + r\mu_1 + \frac{r^2\mu_2}{2!} + \ldots = M(r) = e^{mr+\frac{1}{2}sr^2},$$

where μ_1, μ_2, etc. are the moments of x about zero. Let $M = e^{m+\frac{1}{2}s}$, $l = e^s$. Then $\mu_r' = M^r l^{\frac{1}{2}r(r-1)}$. Hence the cumulants of X are

$$\begin{aligned}
\kappa_1 &= M,\\
\kappa_2 &= M^2(l-1),\\
\kappa_3 &= M^3(l-1)^2(l+2),\\
\kappa_4 &= M^4(l-1)^3(l^3+3l^2+6l+6),\\
\kappa_5 &= M^5(l-1)^4(l^6+4l^5+10l^4+20l^3+30l^2+36l+24),\\
\kappa_6 &= M^6(l-1)^5(l^{10}+5l^9+15l^8+35l^7+70l^6+120l^5+180l^4\\
&\qquad +240l^3+270l^2+240l+120),
\end{aligned} \tag{16}$$

or, if $l-1$ is a small quantity, q, approximating to s,

$$\begin{aligned}
\kappa_1 &= M,\\
\kappa_2 &= M^2q,\\
\kappa_3 &= M^3q^2(3+q),\\
\kappa_4 &= M^4q^3(16+15q+6q^2+q^3),\\
\kappa_5 &= M^5q^4(125+222q+205q^2+120q^3+45q^4+10q^5+q^6),\\
\kappa_6 &= M^6q^5(1296+3660q+5700q^2+5165q^3+4945q^4+2997q^5\\
&\qquad +1365q^6+455q^7+105q^8+15q^9+q^{10}).
\end{aligned}$$

Hence

$$c^2 = q,$$

$$\left.\begin{aligned}
\gamma_1 &= 3c+c^3,\\
\gamma_2 &= 16c^2+15c^4+\ldots,\\
\gamma_3 &= 125c^3+222c^5+\ldots,\\
\gamma_4 &= 1296c^4+3660c^6+\ldots.
\end{aligned}\right\} \tag{17}$$

The first terms represent the limiting values of equations (3) and (9). The leading term of γ_r up to γ_4 at least, is $(r+2)^r$. Thus for a given coefficient of variation, γ_1 is 50 % greater than in the case of the cubed normal variate, γ_2 is 157 % greater.

10. Biological applications

Wilson and Hilferty (1931) showed that the cube root of χ^2 is almost normally distributed when n exceeds 2. Haldane (1938) gave the value of the cumulants in this case, and showed that, for large values of n, $(13\chi^2 - n)^{\frac{5}{13}}$ is even more nearly normally distributed. It is interesting to compare the cumulants of the cubed-normal distribution [equations (2)] with those of the χ^2 distribution. The latter are $\kappa_1 = n$, $\kappa_2 = 2n$, $\kappa_3 = 8n$, $\kappa_4 = 48n$, etc., so that $c = \sqrt{\frac{2}{n}}$, $\gamma_1 = 2c$, $\gamma_2 = 6c^2$. Thus if we compare the χ^2 distribution with a cubed-normal distribution of the same coefficient of variation, we find that they are equally asymmetrical, but that the χ^2 distribution has a γ_2 which is $\frac{27}{28}$th that of the cubed normal. Hence the same transformation will nearly abolish both κ_3 and κ_4 for both distributions.

The success of Wilson and Hilferty's transformation suggests strongly that we may use equations (1), (5) or (3) for the approximate normalization of moderately skew variates. This is an urgent problem in several applications of statistics to biology (Haldane, 1939). We evaluate a number whose mean and standard deviation in the case of random sampling are known. We desire to know whether it differs significantly from the mean. But the sampling distribution is found to be skew. If we can approximately normalize it, our tests of significance become far sharper. This problem is taken up in detail elsewhere. If $\gamma_1^2 > \frac{9}{16}\gamma_2$, and both are small, so that k^2 can be neglected, we can find m, n and k so that $(m + k^{\frac{1}{2}}x)^n$ has a given γ_1 and γ_2. By equations (3) and (4),

$$n = 1 + \frac{4\gamma_1^2}{16\gamma_1^2 - 9\gamma_2}, \quad k = \frac{\gamma_1^2}{9(n-1)^2}, \quad m^{2n} = \frac{\kappa_2}{n^2 k}.$$

If, however, γ_1 is not small, it is necessary to take several terms of equations (3).

If in an observed distribution of weights or volumes, the estimate of γ_1/c, or of $\kappa_1\kappa_3/\kappa_2^2$ is approximately 2, it will be reasonable to try whether γ_2/c^2 or $\kappa_1^2\kappa_4/\kappa_2^2$ approximates to $\frac{56}{9}$, and if so to try to fit a normal distribution to the cube roots of the variate. It will be seen that this does not imply that all the objects considered are of the same shape. On the contrary, such a distribution is to be expected whenever three mutually perpendicular measurements are normally distributed, provided that their coefficients of variation and correlation are not very different. And even the greatest differences in the latter, provided they are not negative, will have little effect.

Unfortunately, reliable estimates of γ_1 and γ_2 can only be obtained from samples of the order of 1000 or more. Rendel (in a paper to be published shortly) obtained the following estimates for the cumulants of the weight distribution of 1202 viable duck eggs, corrected for grouping. The unit is a gram:

$$k_1 = 73{\cdot}294, \quad k_2 = 42{\cdot}69, \quad k_3 = +118{\cdot}658, \quad k_4 = +937{\cdot}21.$$

Hence the estimates of c, γ_1 and γ_2 are

$$c = 0{\cdot}0887 \pm 0{\cdot}0018, \quad g_1 = +0{\cdot}432 \pm 0{\cdot}071, \quad g_2 = +0{\cdot}524 \pm 0{\cdot}142.$$

The standard errors are those appropriate to a normal distribution, but the true values cannot be very different. It is clear that g_1 and g_2 are significantly greater than the values of 0·27 and 0·13 which we should expect were the logarithms of weights normally distributed. They differ still more from what would be expected were the cube roots of weights normally distributed, or on any other hypothesis leading to a similar distribution.

Pearl (1905) lists the moments of the distributions of brain weights of eight European populations of 197 to 529 individuals. The various estimates of c are all close to 0·080, ranging from 0·074 to 0·083. Those of γ_1 range from $+0{\cdot}11$ to $+0{\cdot}40$, those of γ_2 from $-0{\cdot}30$ to $+1{\cdot}5$. The weighted means are

$$c = 0{\cdot}07965 \pm 0{\cdot}00106, \quad g_1 = +0{\cdot}2306 \pm 0{\cdot}0461, \quad g_2 = +0{\cdot}2661 \pm 0{\cdot}0922.$$

If the cube roots are normally distributed, we should expect $g_1 = +0{\cdot}16$, $g_2 = +0{\cdot}037$; if their logarithms are normally distributed, we should expect $g_1 = +0{\cdot}24$, $g_2 = +0{\cdot}10$. The latter distribution gives the better fit, but the first is not impossible.

Sinnott (1937) gives a graph of the distribution of the weights of squash fruits in an F_2, which is positively skew. He shows that a graph of the distribution of their logarithms, though negatively skew, is more symmetrical. There is a suggestion that a graph of the distribution of their cube roots would be even more so. Unfortunately the actual figures are not given, and since curve-fitting by eye is notoriously uncertain, no more can be said. It is much to be desired that, when the full data are not given, estimates of the first four moments or cumulants should be published.

11. SUMMARY

The first four moments of a number of powers and products of normal variates are calculated, with special reference to the probable distributions of weights, volumes, or areas of organs and organisms. In each case the first two measures (γ_1 and γ_2 or β_1 and β_2) of deviation from normality are obtained in terms of the coefficients of variation. The expressions obtained are almost independent of the correlation between the linear measurements, provided the coefficient of variation of the latter are approximately equal. The distribution found is perhaps applicable to data on brain weights, but not to data on ducks' eggs.

REFERENCES

CRAIG, C. C. (1936). On the frequency function of xy. *Ann. Math. Statist.* **7**, 1.

HALDANE, J. B. S. (1938). The approximate normalization of a class of frequency distributions. *Biometrika*, **29**, 392–404.

—— (1939). New data on partial sex-linkage in man. *Proc. Int. Genet. Congr.* p. 137.

MACALISTER, D. (1879). The law of the geometric mean. *Proc. Roy. Soc.* **24**, 367.

PEARL, R. (1905). Biometrical studies on man. I. Variation and correlation in brain-weight. *Biometrika*, **4**, 13–104.

PEARSON, K. (1905). Das Fehlergesetz und seine Verallgemeinerungen durch Fechner und Pearson. A rejoinder. *Biometrika*, **4**, 169–212.

SINNOTT, E. W. (1937). The relation of gene to character in quantitative inheritance. *Proc. Nat. Acad. Sci., Wash.*, **23**, 224–7.

WILSON, E. B. & HILFERTY, M. M. (1931). The distribution of Chi-square. *Proc. Nat. Acad. Sci., Wash.*, **17**, 684.

THE TRANSFORMATION OF DATA FROM ENTOMOLOGICAL FIELD EXPERIMENTS SO THAT THE ANALYSIS OF VARIANCE BECOMES APPLICABLE†

By GEOFFREY BEALL

Dominion Entomological Laboratory, Chatham, Ontario, Canada

1. Introductory

The present paper deals with experiments on the control of insects in the field. In such experimental work the problem to be investigated is whether more insects survive on plots which have been subjected to one treatment than on plots subjected to another. It will be shown in the present paper that the numbers of insects found per plot must vary in such a way that one cannot, strictly, subject the results to the analysis of variance, and it is proposed to find how the data may be transformed so that analysis of variance becomes applicable. Such transformation has been discussed by Bartlett (1936*a, b*) in connexion with entomological experiments, and by Tippett (1934) in connexion with industrial experiments.

2. Experimental results considered

The data used in the following work are results from seven insecticidal experiments arranged by the author at Chatham, Ontario. The work was carried out with replicated blocks containing plots subjected to treatments of which the assignment was random. This procedure, normal in agronomic work, was supplemented by one repetition of each treatment within a block. The assignment of the repetition of a treatment was independent of the first for that treatment, except that, of course, the same plot could not be chosen twice. This repetition was carried out to obtain estimates of variability within blocks. In these experiments complete counts were not made but random sampling was employed. Experiments on *Pyrausta nubilalis* Hubn., reported by Beall *et al.* (1939), for which results are shown in Tables 1 and 2, were made on one area at two different periods, whereas experiments on *Leptinotarsa decemlineata* Say, for which results are indicated in Tables 3 and 4, were carried out on contiguous areas at the same time. Three similar experiments were carried out in one place on the tobacco hornworm, *Phlegethontius quinquemaculata* Haw., for which the data are shown in Tables 5–7. Reference is also made to the data from a uniformity trial on insects of Beall (1939).

† Publication No. 2101, Division of Entomology, Science Service, Department of Agriculture, Ottawa, Canada.

Table 1. *Numbers of an insect,* Pyrausta nubilalis, *per plot. Experiment I*

Treatment	Block 1	2	3	4	5	6	7	8	9	10
1	15	23	21	31	22	14	18	11	21	34
1	27	20	23	33	34	27	17	13	20	26
2	19	12	34	16	20	10	24	23	14	13
2	11	28	37	16	26	18	19	13	10	9
3	16	15	22	25	13	21	18	38	27	10
3	19	16	18	21	19	24	21	18	12	18
4	14	23	10	19	17	18	7	18	8	17
4	34	21	9	34	19	9	15	16	12	12
5	16	16	19	26	15	11	18	23	27	6
5	23	12	12	10	12	17	13	21	12	9
6	12	14	17	10	14	24	24	17	7	3
6	15	16	15	28	13	22	11	7	5	4
7	43	28	35	36	50	69	62	63	42	40
7	47	81	30	69	35	29	71	47	50	43

Table 2. *Numbers of an insect,* Pyrausta nubilalis, *per plot. Experiment II*

Treatment	Block 1	2	3	4	5	6	7	8	9	10
1	32	38	27	7	13	14	26	25	22	30
1	18	40	39	12	19	26	30	19	18	28
2	6	23	8	4	3	18	26	27	17	19
2	9	14	20	13	15	14	15	19	19	10
3	10	21	25	10	13	20	33	48	28	27
3	4	21	26	4	9	14	30	18	27	18
4	2	17	11	3	10	10	26	13	22	17
4	24	13	13	10	6	14	28	11	34	7
5	13	2	5	0	18	10	33	23	20	34
5	17	22	23	8	14	16	26	22	15	34
6	13	10	21	4	10	8	17	15	13	16
6	17	9	29	5	18	15	19	16	27	23
7	37	58	28	11	24	44	30	44	56	45
7	44	71	55	20	26	27	43	52	39	58

Table 3. *Numbers of an insect,* Leptinotarsa decemlineata, *per plot. Experiment III*

Treatment	Block						
	1	2	3	4	5	6	7
1	305	391	420	355	287	175	454
1	207	364	639	527	293	248	397
2	97	49	21	12	3	10	10
2	93	51	25	37	4	12	1
3	270	105	341	469	82	57	221
3	153	190	348	212	100	285	309
4	7	42	34	8	1	10	4
4	12	2	22	4	1	3	3

Table 4. *Numbers of an insect,* Leptinotarsa decemlineata, *per plot. Experiment IV*

Treatment	Block					
	1	2	3	4	5	6
1	253	145	309	665	99	93
1	239	265	166	230	302	237
2	16	13	74	110	14	5
2	95	54	159	108	14	13
3	18	130	165	137	153	78
3	40	137	118	142	239	63
4	2	0	22	6	129	3
4	2	1	31	8	9	8

Table 5. *Numbers of an insect,* Phlegethontius quinquemaculata, *per plot. Experiment V*

Treatment	Block					
	1	2	3	4	5	6
1	6	5	6	13	6	11
1	4	15	13	6	10	15
2	0	1	1	0	1	1
2	2	2	1	4	1	1
3	15	17	22	28	8	16
3	12	22	16	11	13	25

Table 6. *Numbers of an insect,* Phlegethontius quinquemaculata, *per plot. Experiment VI*

Treatment	Block 1	2	3	4	5	6
1	12	13	9	4	11	4
1	13	9	5	7	5	10
2	13	6	8	1	5	7
2	20	9	9	4	12	7
3	7	9	5	4	8	9
3	7	9	4	7	3	2
4	1	1	0	1	4	3
4	2	2	2	1	4	5
5	13	7	12	3	6	11
5	11	5	4	1	9	8
6	7	6	8	9	6	5
6	8	10	2	4	4	12

Table 7. *Numbers of an insect,* Phlegethontius quinquemaculata, *per plot. Experiment VII*

Treatment	Block 1	2	3	4	5	6
1	10	20	14	10	17	14
1	7	14	12	23	20	13
2	11	21	16	17	19	7
2	17	11	14	17	21	13
3	0	7	3	2	3	1
3	1	2	1	1	0	4
4	3	12	4	5	5	2
4	5	6	3	5	5	4
5	3	3	3	1	3	6
5	5	5	6	1	2	4
6	11	15	15	13	26	24
6	9	22	16	10	26	13

3. The relationship between the standard deviation and the mean in the experimental data

If x is the number of insects on one of a group of small contiguous areas, say plots, within a larger area, say a block, let the expectation of x over all these plots be M and the standard deviation be σ; then over a number of the larger

areas, when the insects are distributed in a completely random fashion, from the Poisson distribution,

$$\sigma^2 = M. \tag{1}$$

As is discussed by 'Student' (1919) one cannot, however, anticipate that (1) will be satisfied when organisms occur in groups, as, say, when insects come from masses of eggs, or when there is a change in expectation from plot to plot within a block. Generally, σ^2 will tend to be greater than M and we can only say

$$\sigma^2 = f(M). \tag{2}$$

The form of $f(M)$, in (2), must be considered carefully, since it bears on the form of the transformation which may be developed to make the standard deviation independent of the mean.

In dealing with (2), Bartlett (1936*a*) started by supposing that, approximately,

$$\sigma^2 = KM, \tag{3}$$

where K is a constant. Generally, in field data, however, the relationship between σ^2 and M, or of their respective estimates, s^2 and $\bar{x}$, does not, as in Fig. 1, appear to be linear; rather, the departure of s^2 from $\bar{x}$ becomes disproportionately great as $\bar{x}$ increases. This relationship between departures and the magnitude of the mean has been discussed by Clapham (1936) in connexion with data on the distribution of organisms differing from insects as much as flowering plants, and he showed that only those distributions with very low mean have the squared standard deviation close to the mean.

Our discussion above on the shortcomings of (3) suggests the conclusion that

$$\sigma^2 - M \propto M \tag{4}$$

is generally untrue. We propose to consider the possibility that the curvilinearity of (2) might be better met by supposing that

$$\sigma^2 - M \propto M^2. \tag{5}$$

Equation (5) leads to

$$\sigma^2 = M + kM^2, \tag{6}$$

where k is a constant. It will be noticed that

$$k = (\sigma^2 - M)\, M^{-2} \tag{7}$$

is the Charlier coefficient of disturbance from a Poisson distribution. This coefficient was employed by Beall (1935).

It is possible to consider the suitability of (3), as compared with (6), by finding how, respectively, they fit observations on s^2 and $\bar{x}$. To fit exactly is difficult, and it was found necessary to fall back on an empirical determination of K and of k; thus, if there are a number of pairs of estimates, $\bar{x}$ and s^2, from (3) and (6) we estimate

$$K = \Sigma s^2 / \Sigma \bar{x}, \tag{8}$$

$$k = (\Sigma s^2 - \Sigma \bar{x}) / \Sigma \bar{x}^2, \tag{9}$$

where Σ represents the summation over all pairs.

Since in the work presented in § 2, $\bar{x}$ and s^2, being based on only two observations, are highly variable, these experiments do not show clearly the suitability of (3) and (6). Accordingly, reference is made instead to the data from the uniformity trial on *Leptinotarsa decemlineata* Say of Beall (1939). When the mean and standard deviation of 144 sampling units within each of 16 areas were considered, the estimates from (8) and (9) were $K = 2{\cdot}405$ and $k = 0{\cdot}2548$. For these values from (1), (3) and (6), curves, described as lines 1, 2 and 3 respectively, are plotted in Fig. 1, the observed values of mean and squared standard deviation

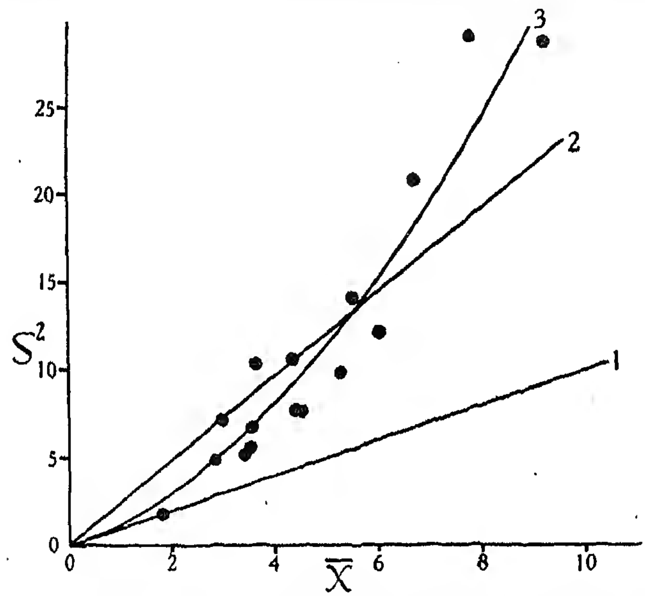


Fig. 1. The squared standard deviation plotted against the mean for 144 small areas within each of 16 large areas; line 1 is from equation (1), line 2 from (3) and line 3 from (6). The counts had been made on *Leptinotarsa decemlineata* Say.

are also shown. In the cases where the mean is near unity the departure of the squared standard deviation from the mean, i.e. from line 1, appears to be trivial, but as the mean increases the departure becomes more marked. It can be seen that the observations lie more snugly about line 3 from (6) than about line 2 from (3). Generally, for the data from field studies the same effect has been observed. Such results suggest that (6) may be generally a better approximation to the form of $f(M)$ than (3) and make it preferable to proceed with the analysis of data from the assumption (6).

4. The transformations of field data

Fig. 1 shows clearly how, within an area, the variability of the numbers of insects on sub-areas is related to the mean number of insects per sub-area. This relationship will make invalid the use of the analysis of variance on experimental

results involving counts on insects, since the expectation of the variance should be the same for all plots. To overcome this invalidity, Bartlett (1936*a*) suggested transforming the observations, x, from the basis of (3). The transformation found was $x^{\frac{1}{2}}$, which Bartlett modified to $(x+\frac{1}{2})^{\frac{1}{2}}$. From § 3 it was seen, however, that for field data the relationship between standard deviation and mean may be represented better by equation (6) than by (3), and, since the form of the transformation depends on the form of $f(M)$, a fresh transformation must be sought. A transformation, as is developed in the Appendix to the present paper, is suggested by the method of Tippett (1934), i.e.

$$x' = k^{-\frac{1}{2}} \sinh^{-1} (kx)^{\frac{1}{2}}. \tag{10}$$

An advance note of this transformation was published by Beall (1940). The adequacy of this transformation must be judged from the extent to which it stabilizes variability. In (10), if we express $\sinh^{-1} (kx)^{\frac{1}{2}}$, when $kx < 1$, as a well-known series, we have

$$x' = x^{\frac{1}{2}} - \tfrac{1}{6} kx^{\frac{3}{2}} + \tfrac{3}{40} k^2 x^{\frac{5}{2}} - \tfrac{5}{112} k^3 x^{\frac{7}{2}} + \ldots, \tag{11}$$

where it is obvious that for $k = 0$, $x' = x^{\frac{1}{2}}$. Of course, for large values of kx, x' varies almost as $\log x$, or as the $\log (x+1)$ used by Williams (1937), and so our proposed expansion may be regarded, for practical purposes, as embracing the root and logarithmic transformations.

Table 8 gives the transformation, (10), for a probable range of observations, x, and for k at intervals which will probably be close enough for practical purposes. This table was computed in part by inverse interpolation from the table of hyperbolic functions of the *Smithsonian Mathematical Tables* (Becker & Van Orstrand, 1931), and in part from (12). Should values of x' be required outside those of Table 8, these can conveniently be calculated from

$$x' = k^{-\frac{1}{2}} \log_e \{(kx)^{\frac{1}{2}} + (1+kx)^{\frac{1}{2}}\}. \tag{12}$$

In preparing Table 8 the question arose of whether, instead of dealing with $k^{-\frac{1}{2}} \sinh^{-1} (kx)^{\frac{1}{2}}$, one should not use $k^{-\frac{1}{2}} \sinh^{-1} \{k^{\frac{1}{2}}(x+\frac{1}{2})^{\frac{1}{2}}\}$ in the same way as Bartlett (1936*a*) dealt with the transformation, $(x+\frac{1}{2})^{\frac{1}{2}}$, instead of $x^{\frac{1}{2}}$. This modification was rejected on the basis of results of the transformation, as discussed in § 5, since it was found that the addition of $\frac{1}{2}$ made little difference and did not give, consistently, an improvement.

For field data, in making the transformation (10), it is necessary to estimate the value of k empirically by (9) for which estimates $\bar{x}$, of the mean and s, of the standard deviation, must be found. The most obvious method in practice of making these estimates seems to be to put more than one plot subjected to a given treatment in a block and so to estimate the chance variation of results for a plot within a block. In the present work, as is discussed in § 2, two plots were subjected to a given treatment in each block and this is probably good practice.

Table 8. *The transformation* $x' = k^{-\frac{1}{2}} \sinh^{-1} (kx)^{\frac{1}{2}}$

x \ k	0·00	0·02	0·04	0·06	0·08	0·10	0·15	0·20	0·25	0·30	0·40	0·50	0·60	0·80	1·00
0	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00	0·00
1	1·00	1·00	0·99	0·99	0·99	0·98	0·98	0·97	0·96	0·96	0·94	0·93	0·92	0·90	0·88
2	1·41	1·40	1·40	1·39	1·38	1·37	1·35	1·33	1·32	1·30	1·27	1·25	1·22	1·18	1·15
3	1·73	1·72	1·70	1·68	1·67	1·66	1·62	1·59	1·57	1·54	1·50	1·46	1·42	1·37	1·32
4	2·00	1·97	1·95	1·93	1·91	1·89	1·84	1·80	1·76	1·73	1·67	1·62	1·58	1·50	1·44
5	2·24	2·20	2·17	2·14	2·11	2·08	2·02	1·97	1·92	1·88	1·81	1·75	1·70	1·61	1·54
6	2·45	2·40	2·36	2·32	2·29	2·25	2·18	2·12	2·06	2·01	1·93	1·86	1·80	1·71	1·63
7	2·65	2·59	2·54	2·49	2·45	2·41	2·32	2·25	2·18	2·13	2·04	1·96	1·89	1·78	1·70
8	2·83	2·76	2·70	2·64	2·59	2·54	2·45	2·36	2·29	2·23	2·13	2·04	1·97	1·85	1·76
9	3·00	2·92	2·84	2·78	2·72	2·67	2·56	2·47	2·39	2·32	2·21	2·12	2·04	1·92	1·82
10	3·16	3·07	2·98	2·91	2·85	2·79	2·66	2·56	2·48	2·40	2·28	2·18	2·10	1·97	1·87
11	3·32	3·21	3·11	3·03	2·96	2·89	2·76	2·65	2·56	2·48	2·35	2·25	2·16	2·02	1·91
12	3·46	3·34	3·23	3·14	3·07	3·00	2·85	2·73	2·63	2·55	2·41	2·30	2·21	2·07	1·96
13	3·61	3·47	3·35	3·25	3·17	3·09	2·93	2·81	2·70	2·62	2·47	2·36	2·26	2·11	1·99
14	3·74	3·59	3·46	3·35	3·26	3·18	3·01	2·88	2·77	2·68	2·52	2·40	2·31	2·15	2·03
15	3·87	3·70	3·56	3·45	3·35	3·26	3·08	2·94	2·83	2·73	2·57	2·45	2·35	2·19	2·06
16	4·00	3·81	3·66	3·54	3·43	3·34	3·15	3·01	2·89	2·79	2·62	2·49	2·39	2·22	2·09
17	4·12	3·92	3·76	3·63	3·51	3·42	3·22	3·07	2·94	2·84	2·67	2·53	2·42	2·25	2·12
18	4·24	4·02	3·85	3·71	3·59	3·49	3·28	3·12	2·99	2·88	2·71	2·57	2·46	2·28	2·15
19	4·36	4·12	3·94	3·79	3·67	3·56	3·34	3·18	3·04	2·93	2·75	2·61	2·49	2·31	2·18
20	4·47	4·22	4·02	3·87	3·74	3·62	3·40	3·23	3·09	2·97	2·79	2·64	2·52	2·34	2·20
21	4·58	4·31	4·11	3·94	3·80	3·69	3·46	3·28	3·13	3·01	2·82	2·68	2·56	2·37	2·23
22	4·69	4·40	4·18	4·01	3·87	3·75	3·51	3·32	3·18	3·05	2·86	2·71	2·58	2·39	2·25
23	4·80	4·49	4·26	4·08	3·93	3·81	3·56	3·37	3·22	3·09	2·89	2·74	2·61	2·42	2·27
24	4·90	4·57	4·34	4·15	3·99	3·86	3·61	3·41	3·26	3·13	2·92	2·77	2·64	2·44	2·29
25	5·00	4·66	4·41	4·21	4·05	3·92	3·65	3·46	3·29	3·16	2·95	2·79	2·66	2·46	2·31
26	5·10	4·74	4·48	4·27	4·11	3·97	3·70	3·49	3·33	3·20	2·98	2·82	2·69	2·48	2·33
27	5·20	4·82	4·54	4·33	4·16	4·02	3·74	3·53	3·37	3·23	3·01	2·85	2·71	2·51	2·35
28	5·29	4·89	4·61	4·39	4·22	4·07	3·78	3·57	3·40	3·26	3·04	2·87	2·73	2·53	2·37
29	5·39	4·97	4·67	4·45	4·27	4·12	3·82	3·61	3·43	3·29	3·07	2·89	2·76	2·54	2·39
30	5·48	5·04	4·74	4·51	4·32	4·16	3·86	3·64	3·46	3·32	3·09	2·92	2·78	2·56	2·40
31	5·57	5·11	4·80	4·56	4·37	4·21	3·90	3·67	3·50	3·35	3·12	2·94	2·80	2·58	2·42
32	5·66	5·18	4·86	4·61	4·42	4·25	3·94	3·71	3·53	3·38	3·14	2·96	2·82	2·60	2·43
33	5·74	5·25	4·91	4·66	4·46	4·30	3·98	3·74	3·55	3·40	3·16	2·98	2·84	2·62	2·45
34	5·83	5·32	4·97	4·71	4·51	4·34	4·01	3·77	3·58	3·43	3·19	3·00	2·86	2·63	2·46
35	5·92	5·38	5·03	4·76	4·55	4·38	4·05	3·80	3·61	3·45	3·21	3·02	2·88	2·65	2·48
36	6·00	5·45	5·08	4·81	4·59	4·42	4·08	3·83	3·64	3·48	3·23	3·04	2·89	2·66	2·49
37	6·08	5·51	5·13	4·85	4·64	4·46	4·11	3·86	3·66	3·50	3·25	3·06	2·91	2·68	2·51
38	6·16	5·57	5·18	4·90	4·68	4·49	4·14	3·89	3·69	3·53	3·27	3·08	2·93	2·69	2·52
39	6·24	5·63	5·23	4·94	4·72	4·53	4·17	3·91	3·71	3·55	3·29	3·10	2·94	2·71	2·53

40	6·32	5·69	5·28	4·99	4·76	4·57	4·20	3·94	3·74	3·57	3·31	3·12	2·96	2·72	2·54
41	6·40	5·75	5·33	5·03	4·79	4·60	4·23	3·97	3·76	3·59	3·33	3·13	2·98	2·73	2·56
42	6·48	5·81	5·38	5·07	4·83	4·63	4·26	3·99	3·78	3·61	3·35	3·15	2·99	2·75	2·57
43	6·56	5·86	5·43	5·11	4·87	4·67	4·29	4·02	3·81	3·63	3·37	3·17	3·01	2·76	2·58
44	6·63	5·92	5·47	5·15	4·90	4·70	4·32	4·04	3·83	3·65	3·39	3·18	3·02	2·77	2·59
45	6·71	5·97	5·52	5·19	4·94	4·73	4·35	4·07	3·85	3·67	3·40	3·20	3·03	2·79	2·60
46	6·78	6·03	5·56	5·23	4·97	4·76	4·37	4·09	3·87	3·69	3·42	3·21	3·05	2·80	2·61
47	6·86	6·08	5·61	5·27	5·01	4·80	4·40	4·11	3·89	3·71	3·44	3·23	3·06	2·81	2·62
48	6·93	6·13	5·65	5·30	5·04	4·83	4·42	4·13	3·91	3·73	3·45	3·24	3·08	2·82	2·63
49	7·00	6·18	5·69	5·34	5·07	4·85	4·45	4·16	3·93	3·75	3·47	3·26	3·09	2·83	2·64
50	7·07	6·23	5·73	5·38	5·10	4·88	4·47	4·18	3·95	3·77	3·48	3·27	3·10	2·84	2·65
55	7·42	6·47	5·93	5·55	5·26	5·02	4·59	4·28	4·04	3·85	3·56	3·34	3·16	2·90	2·70
60	7·75	6·70	6·11	5·70	5·39	5·15	4·70	4·37	4·13	3·93	3·62	3·40	3·22	2·94	2·74
65	8·06	6·91	6·28	5·85	5·52	5·27	4·79	4·46	4·20	4·00	3·69	3·45	3·27	2·99	2·78
70	8·37	7·11	6·44	5·98	5·64	5·38	4·88	4·54	4·28	4·07	3·74	3·50	3·32	3·03	2·82
75	8·66	7·30	6·58	6·11	5·76	5·48	4·97	4·61	4·34	4·13	3·80	3·55	3·36	3·07	2·86
80	8·94	7·47	6·73	6·23	5·86	5·57	5·05	4·68	4·41	4·19	3·85	3·60	3·40	3·10	2·89
85	9·22	7·64	6·86	6·34	5·96	5·66	5·13	4·75	4·47	4·24	3·90	3·64	3·44	3·14	2·92
90	9·49	7·80	6·98	6·45	6·06	5·75	5·20	4·81	4·52	4·29	3·94	3·68	3·48	3·17	2·95
95	9·75	7·96	7·10	6·55	6·15	5·83	5·26	4·87	4·57	4·34	3·98	3·72	3·51	3·20	2·97
100	10·00	8·10	7·22	6·65	6·23	5·91	5·33	4·93	4·62	4·39	4·02	3·75	3·54	3·23	3·00
110	10·49	8·38	7·43	6·83	6·39	6·05	5·45	5·03	4·72	4·47	4·10	3·82	3·60	3·28	3·05
120	10·95	8·64	7·63	6·99	6·54	6·18	5·56	5·13	4·80	4·55	4·16	3·88	3·66	3·33	3·09
130	11·40	8·88	7·81	7·15	6·67	6·31	5·66	5·21	4·88	4·62	4·23	3·94	3·71	3·37	3·13
140	11·83	9·10	7·98	7·29	6·80	6·42	5·75	5·30	4·96	4·69	4·29	3·99	3·76	3·42	3·17
150	12·25	9·31	8·14	7·42	6·91	6·53	5·84	5·37	5·02	4·75	4·34	4·04	3·80	3·45	3·20
160	12·65	9·51	8·29	7·55	7·02	6·62	5·92	5·44	5·09	4·81	4·39	4·08	3·84	3·49	3·23
170	13·04	9·70	8·43	7·67	7·13	6·72	6·00	5·51	5·15	4·86	4·44	4·13	3·88	3·52	3·26
180	13·42	9·88	8·57	7·78	7·23	6·81	6·07	5·57	5·20	4·92	4·48	4·17	3·92	3·56	3·29
190	13·78	10·05	8·69	7·88	7·32	6·89	6·14	5·63	5·26	4·96	4·52	4·20	3·95	3·59	3·32
200	14·14	10·21	8·81	7·98	7·41	6·97	6·20	5·69	5·31	5·01	4·57	4·24	3·99	3·61	3·34
210	14·49	10·36	8·93	8·08	7·49	7·04	6·26	5·74	5·36	5·05	4·60	4·27	4·02	3·64	3·37
220	14·83	10·51	9·04	8·17	7·57	7·11	6·32	5·79	5·40	5·10	4·64	4·31	4·05	3·67	3·39
230	15·17	10·65	9·14	8·26	7·65	7·18	6·38	5·84	5·45	5·14	4·68	4·34	4·08	3·69	3·41
240	15·49	10·79	9·25	8·34	7·72	7·25	6·43	5·89	5·49	5·18	4·71	4·37	4·11	3·72	3·43
250	15·81	10·92	9·34	8·42	7·79	7·31	6·49	5·93	5·53	5·21	4·74	4·40	4·13	3·74	3·45
260	16·12	11·05	9·44	8·50	7·86	7·37	6·54	5·98	5·57	5·25	4·77	4·42	4·16	3·76	3·47
270	16·43	11·17	9·53	8·58	7·92	7·43	6·58	6·02	5·61	5·28	4·80	4·45	4·18	3·78	3·49
280	16·73	11·29	9·61	8·65	7·99	7·49	6·63	6·06	5·64	5·32	4·83	4·48	4·20	3·80	3·51
290	17·03	11·40	9·70	8·72	8·05	7·54	6·68	6·10	5·68	5·35	4·86	4·50	4·23	3·82	3·53
300	17·32	11·51	9·78	8·79	8·10	7·60	6·72	6·14	5·71	5·38	4·88	4·53	4·25	3·84	3·55

In the special case where there are two plots for the ith treatment ($i = 1, \ldots, n$) in the jth block ($j = 1, \ldots, N$), and so two observations, x_{ij1} and x_{ij2}, the estimate of the mean will be written $x_{ij.}$ and of the squared standard deviation

$$s_{ij}^2 = \tfrac{1}{2}(x_{ij1} - x_{ij2})^2. \tag{13}$$

Then from (9), we estimate k from

$$k = 2\left\{\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}-x_{ij2})^2 - \sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2})\right\}\left\{\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2})^2\right\}^{-1} \tag{14}$$

and the calculation is very light.

5. Results showing the effect of the transformation on the variability of data

The adequacy of our proposed transformation may be judged in two ways: first, with respect to its effect, which we shall consider in the present section, on the differences between repetitions of a treatment within a block, and secondly, with respect to its effect, which we shall consider in § 6, on the behaviour of the quantities submitted to the analysis of variance.

It is a fundamental assumption in the analysis of variance that the chance variability for each plot shall be, when the effect of block and of treatment are removed, normally distributed with a standard deviation common to all plots, in which situation of course the standard deviation of the chance variability for a given plot is independent of the expectation for that plot. In the data of the present work, where each treatment is repeated in each block, it is possible to examine the estimates of this standard deviation, s_{ij}, and of the expectation, $x_{ij.}$. For a clear graphical illustration of the situation consider Fig. 2, as obtained from the original data of Experiment III on *Leptinotarsa decemlineata*, where s_{ij} is plotted against $x_{ij.}$, and contrast this situation with that obtaining for the corresponding quantities s'_{ij} and $x'_{ij.}$, obtained after transformation ($k = 0{\cdot}08$) in Fig. 3.

In Fig. 2 the points are widely scattered as is natural from a sample of two; nevertheless, it is apparent that for the smallest values of $x_{ij.}$ the values of s_{ij} are correspondingly small and fall in a close group. In Fig. 3 the cluster of observations in the lower left-hand corner of the previous diagram has disappeared, and generally the scatter appears to be independent of $x'_{ij.}$, so that apparently the transformation gave satisfactory results. The nature of the material involved is such that it does not seem possible to examine the relationship under consideration more exactly, nor to summarize exactly the corresponding results for the other treatments; it can only be said that the same type of result appeared although the magnitude of the relationship before transformation depended on the magnitude of the differences between the effects of treatments.

The results shown in Figs. 2 and 3 suggest that the proposed transformation has tended to make the standard deviation independent of the mean, in accordance

with the assumptions underlying the analysis of variance. In using this procedure one actually assumes, more broadly, that a common standard deviation exists, so that the homoscedasticity of observations before and after transformation should be tested. Thus it is assumed that x_{ij1} and x_{ij2} are observations from a normal

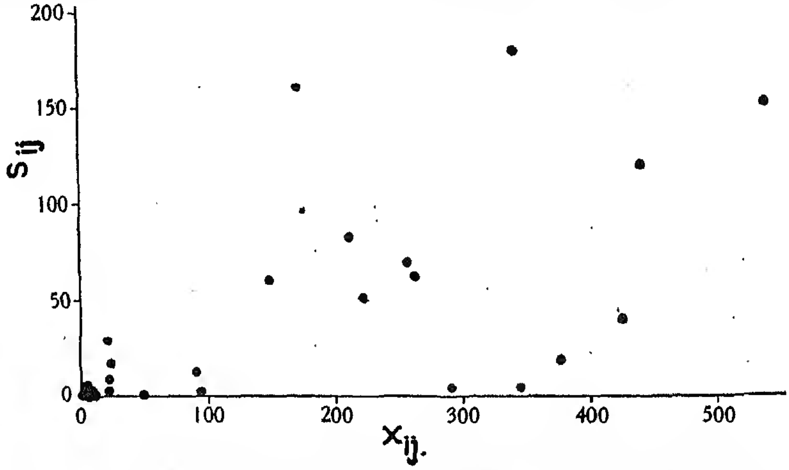


Fig. 2. The standard deviation and mean as estimated from plots by pairs, with untransformed data on *Leptinotarsa decemlineata*.

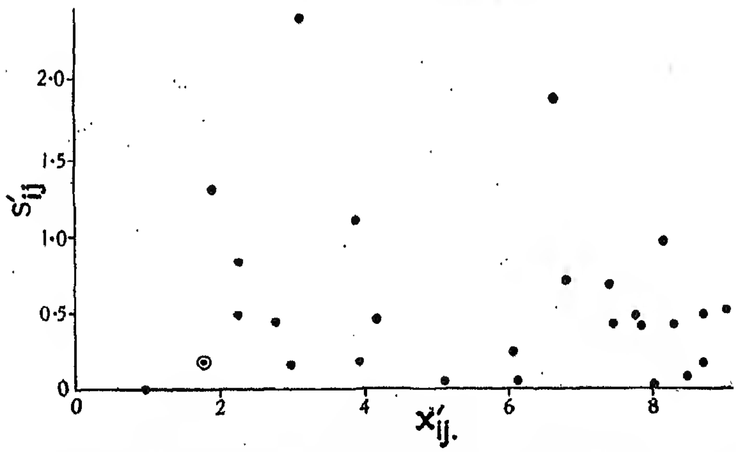


Fig. 3. The standard deviation and mean as estimated from plots by pairs, with the transformed data on *Leptinotarsa decemlineata* Say, i.e. using $x' = k^{-\frac{1}{2}} \sinh^{-1} (kx)^{\frac{1}{2}}$, $(k = 0{\cdot}08)$.

population with a standard deviation, σ, which is independent of i and j. Then $\left\{ \sum_{k=1}^{2} (x_{ijk} - x_{ij.})^2 \right\} \frac{1}{\sigma^2}$ is distributed as χ^2 with one degree of freedom.† Accordingly,

† In the L_1 test, discussed by Nayer (1936), this case of estimates of standard deviation with one degree of freedom is troublesome since zero values tend to arise when dealing with grouped or integral observations. When this is the case L_1, which is the ratio of an arithmetic to a geometric mean of sums of squares, cannot be calculated. The present treatment may therefore have a wider application.

$y_{ij} = (x_{ij1} - x_{ij2})/\sqrt{2}\sigma$ should be distributed normally with unit standard deviation for all i and j. In order to test the hypothesis of normality with unit standard deviation it is only necessary to test for leptokurtosis; for the distribution must be symmetrical since the sign of differences, and therefore of y_{ij}, is a matter of chance. Since the number of items involved will almost certainly be < 100, and since the population mean is zero, the w_n criterion of Geary (1935) will provide an appropriate test. In using this criterion we must find the ratio of the mean deviation to the standard deviation, i.e.

$$w_n = \left\{\sum_{i=1}^{n}\sum_{j=1}^{N} |x_{ij1} - x_{ij2}|\right\}\left\{nN\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1} - x_{ij2})^2\right\}^{-\frac{1}{2}}. \tag{15}$$

Of course, values of w_n may be calculated for transformed data by substitution of x'_{ijk} for x_{ijk}.

Table 9. *The w_n test on the homoscedasticity of counts by plots within a block for six field experiments*

Experiment	nN	Upper 5% limit	Lower 5% limit	Untransformed		Transformed (Bartlett)		Value of k		Transformed (Beall)	
				w_n	Departure by S.D.	w_n	Departure by S.D.	Estimated	Employed	w_n	Departure by S.D.
I	70	0·841	0·757	0·6659	−5·33	0·7525	−1·91	0·078	0·08	0·7846	−0·64
II	70	0·841	0·757	0·7885	−0·49	0·7807	−0·79	0·046	0·04	0·7499	−2·01
III	28	0·866	0·737	0·5973	−5·33	0·6431	−4·15	0·084	0·08	0·6838	−3·11
IV	24	0·872	0·732	0·5692	−5·67	0·6948	−2·67	0·285	0·30	0·7370	−1·66
V	18	0·881	0·728	0·7554	−1·16	0·8155	+0·13	0·082	0·08	0·7823	−0·58
VI	36	0·857	0·745	0·7959	−0·22	0·8166	+0·38	0·019	0·02	0·8130	+0·28

Values of w_n from the untransformed observations and from the transformed observations, both following Bartlett (i.e. the transformation $(x+\frac{1}{2})^{\frac{1}{2}}$) and following the line suggested in the present paper, are shown in Table 9 for the field data of Tables 1–6. For the second transformation the values of k as calculated from (14) are shown as well as the nearest value of k entered in Table 8. For each experiment the value of nN and also the 0·05 limits of probability, from Geary (1935), are shown. There are also shown the departures of observed w_n from the expected value in terms of the standard deviation, a useful criterion since the distribution of w_n is almost normal. From Table 9 it can be seen that out of the three experiments in which w_n fell beyond the lower 5% limit of probability for the untransformed data and the data transformed as $(x+\frac{1}{2})^{\frac{1}{2}}$, in only one experiment did w_n fall so with the final transformation. The results for Experiment II, in which w_n is decreased by the transformation, are peculiar. Consideration of the departures from the mean in terms of the standard deviation indicates more clearly the improvement effected by each transformation

and how the transformation suggested in the present work secures an improvement of the same, but more marked, character than that secured from the transformation of Bartlett. The results suggest that while homoscedasticity may not be attained always, it will be approached by means of the proposed transformation.

6. THE EFFECT OF THE TRANSFORMATION ON THE ANALYSIS OF VARIANCE

As was indicated at the beginning of § 5, our proposed transformation besides making the variability within a block for a repeated treatment the same for all treatments and blocks, should also provide quantities satisfying the assumptions underlying the analysis of variance. Since it is not quite clear how, in so far as the transformation is satisfactory in the first way, it will necessarily be satisfactory in the second, it will be well to consider directly the suitability of our transformed values for the analysis of variance.

In the application of the analysis of variance one would deal with $x_{ij.}$ rather than with x_{ijk} and suppose that

$$x_{ij.} = A + B_i + C_j + D_{ij}, \tag{16}$$

where A is a contribution from the general level of population on the experimental area, B_i the contribution of the ith treatment and C_j the contribution of the jth block. The remainder term, D_{ij}, is called the interaction of treatments and blocks. Of course, the present discussion on the untransformed values, $x_{ij.}$, holds for the transformed values, $x'_{ij.} = \frac{1}{2}(x'_{ij1} + x'_{ij2})$ when the appropriate symbols, A', B'_i, C'_j and D'_{ij} are used.

In material satisfying the conditions underlying the analysis of variance, for the observations under each treatment, the calculated squared standard deviation is

$$s_i^2 = \frac{1}{N-1} \sum_{j=1}^{N} (x_{ij.} - x_{i..})^2. \tag{17}$$

Following the argument of the analysis of variance, $x_{ij.} - x_{i..}$, of which the mean is 0, is an estimate of $C_j + D_{ij}$, in which the two terms are independent; hence the expectation of s_i^2 is

$$\sigma_i^2 = \sigma_{C_j}^2 + \sigma_{D_{ij}}^2, \tag{18}$$

where σ_{Cj} and σ_{Dij} are the standard deviations of the parameters, C_j and D_{ij}, respectively, and are independent of treatment. Accordingly, s_i should be independent of treatment and distributed as an estimate of σ_i, having $N-1$ degrees of freedom. Conversely, if $x_{ij.}$ cannot be built of the independent terms of (16), then the various values of s_i will not be distributed as estimates of a single standard deviation. The hypothesis that the values of s_i in any one experiment are estimates of one quantity may be tested.†

† From correspondence with Dr R. W. B. Jackson, the writer has learned that he had arrived independently at the same test.

The results of the tests on the homogeneity of the values, s_i, within the six experiments treated in the present paper are presented in Table 10, where the value of the L_1 criterion is shown for the original data and for the transformed values together with the appropriate 0·05 and 0·01 levels of probability (Nayer, 1936). From Table 10 it can be seen that of the values of L_1 obtained from the original data, all but one are near or beyond the 0·05 level of significance, but that after transformation all are moved in to less significant values. Accordingly, the values of s_i, when calculated from the original data, appear heterogeneous but the corresponding values obtained after the transformation appear homogeneous. Thus it is more probable that the analysis of variance is applicable to the transformed data than to the untransformed.

Table 10. *The homogeneity, as measured by the criterion L_1, of the estimates s_i for various values of i before and after transformation*

	Experiment					
	1	2	3	4	5	6
L_1 before transformation	0·867	0·833	0·325	0·657	0·344	0·680
L_1 after transformation	0·864	0·941	0·766	0·688	0·813	0·730
1% limit	0·757	0·757	0·604	0·542	0·514	0·583
5% limit	0·812	0·812	0·707	0·656	0·648	0·673

In Table 10 we have tested the homogeneity of the estimates, s_i, as in § 5 we tested the homogeneity of s_{ij}, that is without reference to the values of the associated means. In view of our original assumptions we are, however, interested in the possibility that the standard deviations, as calculated, might show every sign of being estimates of a common standard deviation and yet be dependent on the associated means. Accordingly, we have investigated such dependence roughly by fitting by least squares a first order regression of s_i on $x_{i..}$. From this fitting we record the sign of the regression as follows:

	Experiment					
	1	2	3	4	5	6
Before transformation	+*	+**	+*	+**	+	+
After transformation	−	+	−	−*	+	+

By a single asterisk we have indicated cases where the reduction in variability effected by the regression passed the 5 % probability limit and by a double asterisk where it passed the 1 % limit. Several points may be noted. (1) In two cases (Experiments 2 and 6) after transformation the residual sum of squares about the regression was greater than the reduction in squares due to the regression, whereas it was consistently less before transformation. (2) As can be seen above, the regression generally did not effect a significant reduction in variability after transformation but did before (the small number of degrees of freedom made high significance difficult of attainment). (3) After transformation the sign of the regression seemed to be a chance matter, whereas before transformation it was consistently positive. These results suggest that the transformation proposed did tend to make the variability within a given treatment independent of the mean for that treatment.

7. THE EFFECT OF TRANSFORMATION UPON THE CONCLUSIONS FROM THE ANALYSIS OF VARIANCE

It has been shown in §§ 5 and 6 that the analysis of variance can be made on entomological data when a suitable transformation has been effected. It is of practical interest to see what numerical effect such transformation will have upon tests on the significance of, say, the effect of treatment and the significance of differences for treatments.

First, consider the numerical results to be obtained from the analysis of variance (1) without and (2) with transformation. Thus the mean square ascribable to blocks, treatments and their interaction is shown in Table 11, for six experiments of which the data are given in § 2; parallel results are presented for untransformed observations and for observations transformed by (10) with the values of k from Table 9. To facilitate the comparison of the results, the mean square for blocks and for treatments is expressed in terms of the estimate for interaction, as the F of Snedecor (1934), and presented in each case. The transformation of the data has modified the conclusions to be drawn from the analysis of variance in Table 11, in that there are considerable changes in the criterion, F, for treatments or for blocks. In the examples shown the effect of treatments was highly significant in all cases and so the changes introduced by transformation did not alter the conclusions, as would have been the case for less definite effects.

Consider next the effect of transformation on the significance of differences between the means for treatments as tested by the criterion, t, calculated with such estimates of mean square as the interaction of Table 11. For illustration, values of t, from the data on *Leptinotarsa decemlineata* (Experiment III), are shown in Table 12 for each possible comparison of treaments when untransformed data are used, when the transformation, $(x+\frac{1}{2})^{\frac{1}{2}}$, as suggested by Bartlett (1936*a*) is used, and when the transformation, $k^{-\frac{1}{2}}\sinh^{-1}(kx)^{\frac{1}{2}}$, as suggested in the present paper is used. In order that the influence of the level of population under each

Table 11. *The analysis of variance of untransformed and transformed data in six experiments*

Variation	Degrees of freedom	Untransformed data: Mean square	Untransformed data: F	Transformed data: Mean square	Transformed data: F
Experiment I. *P. nubilalis*					
Between blocks	9	92·8	1·21	0·565	1·66
Between treatments	6	2,839·0	36·95	7·51	22·03
Interaction	54	76·8	—	0·341	—
Experiment II. *P. nubilalis*					
Between blocks	9	577·0	6·72	5·37	10·55
Between treatments	6	1,721·0	20·04	8·69	17·07
Interaction	54	85·9	—	0·509	—
Experiment III. *L. decemlineata*					
Between blocks	6	20,172·0	2·26	4·67	3·03
Between treatments	3	390,932·0	43·77	111·5	72·16
Interaction	18	8,931·0	—	1·54	—
Experiment IV. *L. decemlineata*					
Between blocks	5	12,960·0	2·06	2·06	1·96
Between treatments	3	124,054·0	20·40	20·40	19·41
Interaction	15	6,727·0	—	1·05	—
Experiment V. *P. quinquemaculata*					
Between blocks	5	27·4	2·74	0·349	4·19
Between treatments	2	752·0	75·33	19·5	233·77
Interaction	10	9·98	—	0·083	—
Experiment VI. *P. quinquemaculata*					
Between blocks	5	41·7	4·13	1·50	4·12
Between treatments	5	66·2	6·55	3·33	9·14
Interaction	25	10·1	—	0·365	—

Table 12. *The values of t, in the comparison of means, as calculated from the untransformed and the transformed data of Experiment III on* Leptinotarsa decemlineata

Comparison	Means for untransformed data		t without transformation	t from $(x+\frac{1}{2})^{\frac{1}{2}}$	t from $k^{-\frac{1}{2}}\sinh^{-1}(kx)^{\frac{1}{2}}$
$x_{1..}-x_{2..}$	362	30	+9·27**	+11·33**	+9·92**
$x_{1..}-x_{3..}$	362	224	+3·84**	+3·52**	+2·11*
$x_{1..}-x_{4..}$	362	11	+9·82**	+12·89**	+12·46**
$x_{2..}-x_{3..}$	30	224	−5·43**	−7·81**	−7·81**
$x_{2..}-x_{4..}$	30	11	+0·54	+1·56	+2·54*
$x_{3..}-x_{4..}$	224	11	+5·98**	+9·37**	+10·35**

treatment may be judged, there are shown, also in Table 12, the means for the untransformed data. The values of t falling beyond the 0·01 level of significance have been marked with two asterisks and the values beyond the 0·05 level with one. It can be seen that the transformation resulted in a profound alteration in the conclusions. Apparently on account of the dependence of variance on mean in untransformed data, the pooled estimate of variance was originally too low for the treatments which resulted in high populations and too high for the treatments which resulted in low populations. Thus, in the comparison of the first and third treatments, which appeared to have the two highest surviving populations, the value of t calculated from untransformed values was high. In the other extreme case, the comparison between the second and fourth treatments, the value of t, as calculated from untransformed values was very low. It can be seen further, that the first transformation only secured in part the modification in the value of t that was secured by the second transformation.

8. THE PROCEDURE OF TRANSFORMATION IN PRACTICE

The methods which were found applicable in the preceding discussion will now be illustrated in the transformation of the data shown in Table 7 (Experiment VII) on *Phlegethontius quinquemaculata*, of the same type as the experiments previously discussed in the present paper. The steps in the analysis will be set out with the purpose of providing a model for procedure in estimating the constant, k, which will be used to effect a transformation of the data so that the analysis of variance may be made.

Supposing that the experiment has been laid out with a repetition of each treatment in each block, the procedure of estimating k makes it first necessary to find the sum and the absolute difference of each pair of plots subjected to a given treatment in a given plot and then to sum the sums, $\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2})$, and the sums squared, $\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2})^2$, and also the differences squared, $\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}-x_{ij2})^2$, over all such pairs and by substituting the results in (14) to find k. In the case being used for an illustration the two plots subjected to the first treatment in each block gave respectively 10 and 7, 20 and 14, 14 and 12, 10 and 23, 17 and 20, 14 and 13, so that

$$\sum_{j=1}^{N}(x_{1j1}+x_{1j2}) = (10+7)+(20+14)+(14+12)+\ldots = 174.$$

Similarly,

$$\sum_{j=1}^{N}(x_{1j1}+x_{1j2})^2 = (10+7)^2+(20+14)^2+\ldots = 5308$$

and similarly,

$$\sum_{j=1}^{N}(x_{1j1}-x_{1j2})^2 = (10-7)^2+(20-14)^2+\ldots = 228.$$

Of course, in estimating k the summations are not limited to one treatment but must be extended over all in the experiments. If this is done we find

$$\sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2}) = 684, \quad \sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}+x_{ij2})^2 = 19{,}656, \quad \sum_{i=1}^{n}\sum_{j=1}^{N}(x_{ij1}-x_{ij2})^2 = 708.$$

From (14) we estimate $$k = \frac{2(708-684)}{19{,}656} = 0{\cdot}002,$$

and referring to Table 8, p. 250, use $k = 0{\cdot}00$ as the nearest value occurring there. Of course, in this case, the transformation is simply $x^{\frac{1}{2}}$.

Now from the above result it will be possible to replace the observed values of Table 7 with the corresponding transformed values from the first column of Table 8. Thus in Table 7 replace in the first row: 10, 20, 14, 10, 17 and 14, by 3·16, 4·47, 3·74, 3·16, 4·12 and 3·74. With such transformed values we can now proceed to carry out a routine analysis of variance which will be facilitated by working with the sum for each pair of plots in a given block with a given treatment. For example, the final analysis of variance for Experiment VII would be carried out with the values of Table 13.

Table 13. *Transformed and summed values to be used in the analysis of variance for Experiment VII on* P. quinquemaculata

Treatment	Block					
	1	2	3	4	5	6
1	5·81	8·21	7·20	7·96	8·59	7·35
2	7·44	7·90	7·74	8·24	8·94	6·26
3	1·00	4·06	2·73	2·41	1·73	3·00
4	3·97	5·91	3·73	4·48	4·48	3·41
5	3·97	3·97	4·18	2·00	3·14	4·45
6	6·32	8·56	7·87	6·77	10·20	8·51

9. Summary and conclusions

The foregoing work is a study of experimental results from seven field experiments on the control of insects. In such data, the standard deviation of the number of insects per plot varies with the mean. By the transformation, $x' = k^{-\frac{1}{2}}\sinh^{-1}(kx)^{\frac{1}{2}}$, where k is a constant and x an observation, the data were put in a form for which the standard deviation approached a constant independent of the mean. The estimation of the one constant, k, necessary for the transformation was made possible by the design of the experiments with repetition of treatments within blocks. In practice, the transformation gave good results so that

analysis of variance could be made. From the analysis of the transformed data, the results were found to differ markedly from those which would have been obtained from the untransformed data.

ACKNOWLEDGEMENTS

The present work was suggested to the writer by Prof. E. S. Pearson to whom the writer is further indebted for advice as the work progressed. The writer is further beholden to Dr R. W. B. Jackson, to Dr B. L. Welch and to Dr M. S. Bartlett for criticism and advice. Dr G. M. Stirrett and Mr F. A. Stinson very kindly permitted the use of unpublished data from their experiments carried out at Delhi, Ontario, on *Phlegethontius quinquemaculata* Haw.

APPENDIX

As has been said, the transformation of (10) was suggested by the method used by Tippett (1934, p. 61). The procedure is as follows.

It is required to find $x' = f(x)$, such that the standard deviation, $\sigma_{x'}$ of x', shall be approximately constant. Let us write

$$x' = f(M) + f'(M)(x - M) + \ldots, \tag{19}$$

where M is the expectation of x and whence, approximately,

$$(x' - M') = f'(M)(x - M), \tag{20}$$

where M' is the expectation of x'. Hence

$$\sigma_{x'}^2 = \{f'(M)\}^2 \sigma^2, \tag{21}$$

where σ is the standard deviation of the observations, x. Replacing $\sigma_{x'}$ in (21) by a constant, c, as is the purpose of our operation, and substituting for σ from equation (6), p. 247, we have

$$f'(M) = c(M + kM^2)^{-\frac{1}{2}}, \tag{22}$$

where k is, as has been previously discussed, a constant peculiar to our data. Integrating in (22),

$$f(M) = 2ck^{-\frac{1}{2}} \sinh^{-1}(kM)^{\frac{1}{2}}. \tag{23}$$

From (23) the form of the function suggested is $\sinh^{-1}(kx)^{\frac{1}{2}}$, but it is wise instead to use $k^{-\frac{1}{2}} \sinh^{-1}(kx)^{\frac{1}{2}}$, since the transformation then becomes identical, as shown in (11), with the established transformation, $x^{\frac{1}{2}}$, when $k = 0$.

As Tippett (1934) says: 'This derivation is not mathematically sound, and the result is only justified if on application it is found to be satisfactory.' The writer would have hesitated to have used it had it not already led to useful transformations in cases analogous to the present, namely to $x^{\frac{1}{2}}$ where x comes from a Poisson distribution, to $\sin^{-1} p^{\frac{1}{2}}$ where p comes from a binomial distribution and, according to Tippett, to $\tanh^{-1} r$, where r is the correlation coefficient.

REFERENCES

BARTLETT, M. S. (1936*a*). Square root transformation in analysis of variance. *J. Roy. Statist. Soc.* Suppl. **3**, 68–78.

—— (1936*b*). Some notes on insecticide tests in the laboratory and in the field. *J. Roy. Statist. Soc.* Suppl. **3**, 185–94.

BEALL, G. (1935). Study of arthropod populations by the method of sweeping. *Ecology*, **16**, 216–25.

—— (1939). Methods of estimating the population of insects in a field. *Biometrika*, **30**, 422–39.

—— (1940). The transformation of data from entomological field experiments. *Canadian Ent.* **72**, 168.

BEALL, G., STIRRETT, G. M. & CONNERS, I. L. (1939). A field experiment on the control of the European corn borer, *Pyrausta nubilalis* Hubn., by *Beauveria Bassiana* Vuill. II. *Sci. Agric.* **19**, 531–4.

BECKER, G. F. & VAN ORSTRAND, C. E. (1931). Hyperbolic functions (4th reprint). *Smithsonian Mathematical Tables.*

CLAPHAM, A. R. (1936). Over-dispersion in grassland communities and the use of statistical methods in plant ecology. *J. Ecol.* **24**, 232–51.

GEARY, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, **27**, 310–32.

NAYER, P. P. N. (1936). An investigation into the application of Neyman and Pearson's L_1 test, with tables of percentage limits. *Statist. Res. Mem.* **1**, 38–51.

SNEDECOR, G. W. (1934). *Calculation and Interpretation of Analysis of Variance and Covariance.* Pp. 96. Ames, Iowa: Collegiate Press Inc.

'STUDENT' (1919). An explanation of deviations from Poisson's law in practice. *Biometrika*, **12**, 211–15.

TIPPETT, L. H. C. (1934). Statistical methods in textile research. Part 2. Uses of the binomial and Poisson distributions. *Shirley Inst. Mem.* **13**, 35–72.

WILLIAMS, C. B. (1937). The use of logarithms in the interpretation of certain entomological problems. *Ann. Appl. Biol.* **24**, 404–14.

INTERPOLATION FOR FRESH PROBABILITY LEVELS BETWEEN THE STANDARD TABLE LEVELS OF A FUNCTION

By J. B. SIMAIKA

1. The problem

A number of tables of probability functions exist, and more will no doubt before long be available, giving values for a variable x corresponding to a limited number of simple probability levels α. How far is it possible to obtain x rapidly for intermediate values of α?

The variable may be put into standardized form as the ratio of the deviation from the mean to the standard deviation; these two quantities (i.e. the mean and standard deviation) are often easy to obtain, whereas the probability integral may require extensive computation. Denote by u_α the standardized variable so that

$$u_\alpha = \frac{x_\alpha - \text{mean } x}{\text{standard deviation of } x}. \tag{1}$$

The question we shall consider is this: having full and accurate tables relating u_α and α for a standardized normal variable, denoted by U_α, can we use these values as auxiliary in obtaining u_α for any other function tabled only at a few probability levels and, having found an interpolating formula, what is its accuracy? In examining this point we shall compare the accuracy of the method with that of some other methods of deriving intermediate values of u_α.

2. General approach

It will be useful to consider first how far a general theoretical approach will take us. Let the variable x follow a probability law defined only by its cumulants κ_r $(r = 1, 2, \ldots)$; then the first four cumulants of the variable u become $0, 1, \gamma_1, \gamma_2$. It is known that the relation between u_α and α may be written symbolically

$$\alpha = \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{(2\pi)}} \exp\left[-\tfrac{1}{6}\gamma_1 \frac{d^3}{dx^3} + \tfrac{1}{24}\gamma_2 \frac{d^4}{dx^4} + \ldots\right] e^{-\frac{1}{2}x^2}\, dx, \tag{2}$$

while the same relation for a normal variable is

$$\alpha = \int_{-\infty}^{U_\alpha} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx. \tag{3}$$

Using equations (2) and (3) it has been shown by Cornish & Fisher (1937) that u_α can be approximated to by a parabolic curve in U_α and vice versa. If we

assume that κ_r for $r \geqslant 5$ is negligible, this expression, using a fifth degree parabola, is

$$u_\alpha = A + BU_\alpha + CU_\alpha^2 + DU_\alpha^3 + EU_\alpha^4 + FU_\alpha^5, \tag{4}$$

where

$$\begin{aligned}
A &= -\tfrac{1}{6}\gamma_1 - \tfrac{1}{12}\gamma_1\gamma_2 + \tfrac{17}{324}\gamma_1^3, \\
B &= 1 - \tfrac{1}{8}\gamma_2 + \tfrac{5}{36}\gamma_1^2 - \tfrac{29}{384}\gamma_2^2 - \tfrac{107}{288}\gamma_1^2\gamma_2 - \tfrac{1511}{7776}\gamma_1^4, \\
C &= \tfrac{1}{6}\gamma_1 + \tfrac{5}{24}\gamma_1\gamma_2 - \tfrac{53}{324}\gamma_1^3, \\
D &= \tfrac{1}{24}\gamma_2 - \tfrac{1}{18}\gamma_1^2 + \tfrac{1}{16}\gamma_2^2 - \tfrac{103}{288}\gamma_1^2\gamma_2 + \tfrac{211}{972}\gamma_1^4, \\
E &= -\tfrac{1}{24}\gamma_1\gamma_2 + \tfrac{1}{27}\gamma_1^3, \\
F &= -\tfrac{1}{128}\gamma_2^2 + \tfrac{7}{144}\gamma_1^2\gamma_2 - \tfrac{7}{216}\gamma_1^4.
\end{aligned}$$

Now as γ_1 and γ_2 tend to zero, B tends to unity and all other coefficients tend to zero, i.e. the curve of u_α as a function of U_α tends to the diagonal line

$$u_\alpha = U_\alpha.$$

Figs. 1–4 give these curves for different numbers of degrees of freedom for the commonly used statistical variables χ^2, χ, t and v (a transformation of z referred to below), expressed in standardized form.

With regard to the coefficients in (4) it may be remarked that large values of γ_2 do not increase them as much as large values of γ_1. Furthermore, when γ_1 is zero, the coefficients A, C and E vanish and the expression (4) becomes

$$u_\alpha = BU_\alpha + DU_\alpha^3 + FU_\alpha^5,$$

or

$$\frac{u_\alpha}{U_\alpha} = B + DU_\alpha^2 + F\{U_\alpha^2\}^2. \tag{5}$$

These broad results suggest that a good method of interpolation, when both γ_1 and γ_2 exist, is a Lagrange formula through the points

$$(u_{\alpha_i}, U_{\alpha_i}) \quad (i = 1, 2, \ldots).$$

Remembering that $u_{\alpha_i} = (x_{\alpha_i} - \text{mean}\, x)/(\text{S.D. of } x)$, from the practical point of view the interpolation can be carried out more expeditiously and without loss of accuracy by using a Lagrangian formula through

$$(x_{\alpha_i}, U_{\alpha_i}) \quad (i = 1, 2, \ldots).$$

When, however, γ_1 is zero as in the case of the t-distribution it would be better to take the Lagrange formula through the points

$$\left(\frac{u_{\alpha i}}{U_{\alpha i}}, U_{\alpha i}^2\right) \text{ or alternatively through } \left(\frac{x_{\alpha i}}{U_{\alpha_i}}, U_{\alpha i}^2\right) \quad (i = 1, 2, \ldots),$$

the mean x being zero.

The accuracy of the method is likely to depend on the value of γ_1 and γ_2. For example, as seen in Fig. 1, linear interpolation between, say, $U_{0\cdot05}$ and $U_{0\cdot02}$ will be more accurate with $\nu = 18$ than $\nu = 3$, the γ's being smaller in the former case. Again, the curves will be more nearly linear if we take as variable

$$u = (\chi - \text{mean}\,\chi)/\sigma_\chi \quad \text{rather than} \quad u = (\chi^2 - \text{mean}\,\chi^2)/\sigma_{\chi^2}$$

because the γ's for the former are the smaller.

For the practical worker it will often be sufficient to use linear interpolation, i.e. to make use of two tabled probability levels only. For more accurate work

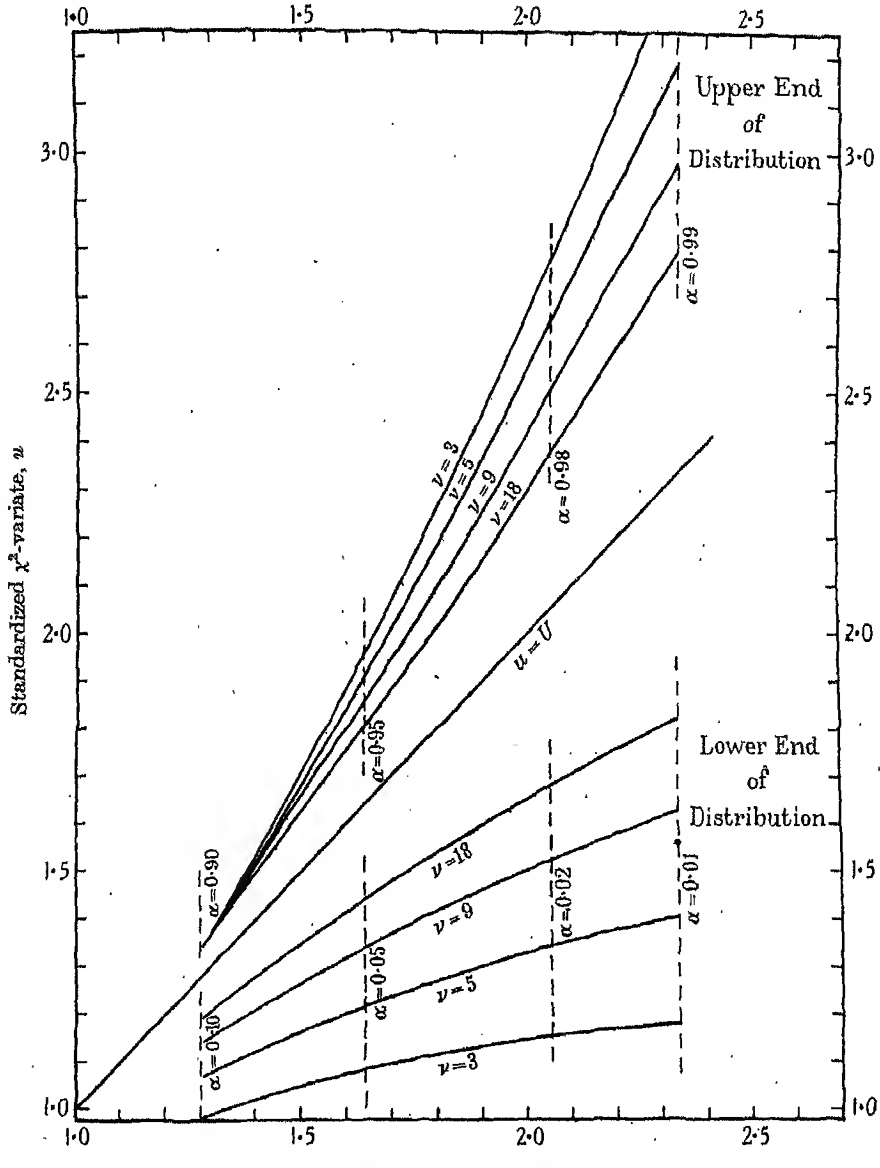


Fig. 1. Relation between $u(\chi^2)$ and U.

three or more levels can be used, but an increase beyond this is not in fact likely to lead to a gain in accuracy which will be worth the labour.

3. Comparison of methods

To compare the accuracy of the interpolation based on the polynomial expansion (4) with that obtained by other possible methods, we have considered the χ^2, t and v (beta)-probability distributions. For each, different methods of interpolation have been devised. In some a transformation of the variable has

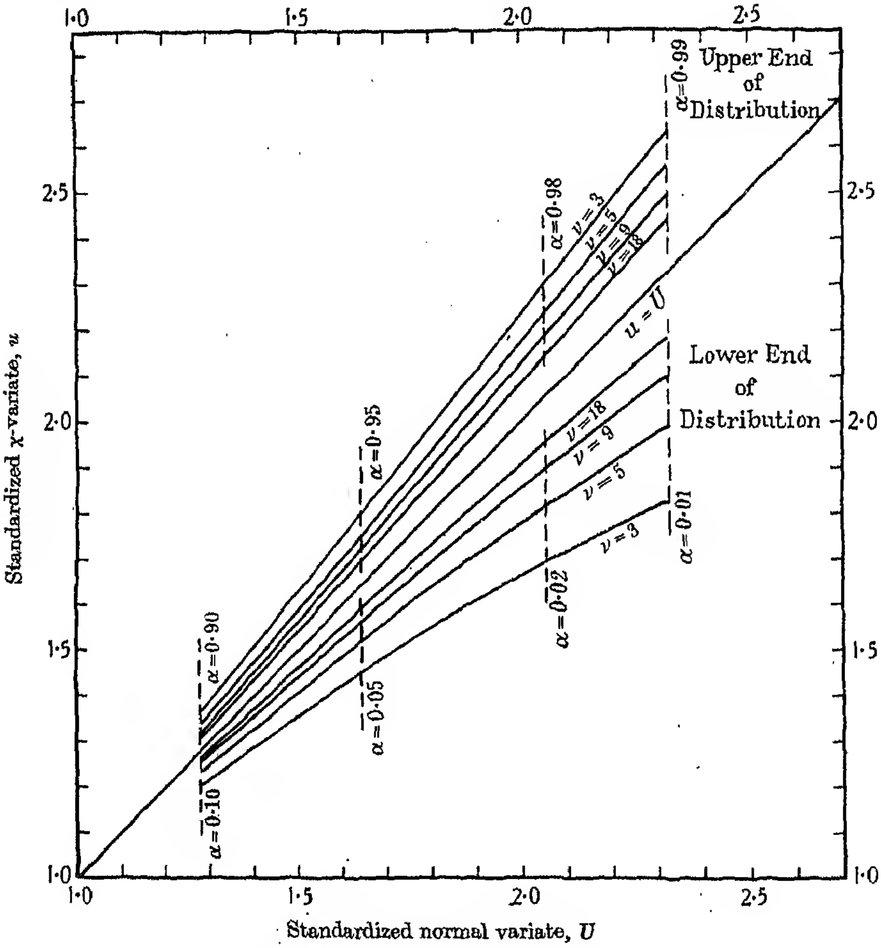


Fig. 2. Relation between $u(\chi)$ and U.

been used, while in others a transformation of the argument or a completely new argument has been considered.

The accuracy of each method in the range covered by the values $\alpha \leqslant 0\cdot10$ and $\alpha \geqslant 0\cdot90$ has been tested in the following way: between any two consecutive probability levels a number—not less than three—of intermediate values have been interpolated. These values were chosen to be those which could be obtained accurately from some other table. The greatest deviations in each interval are

given in Tables 1–3. The methods have been arranged according to the degree of accuracy obtained.

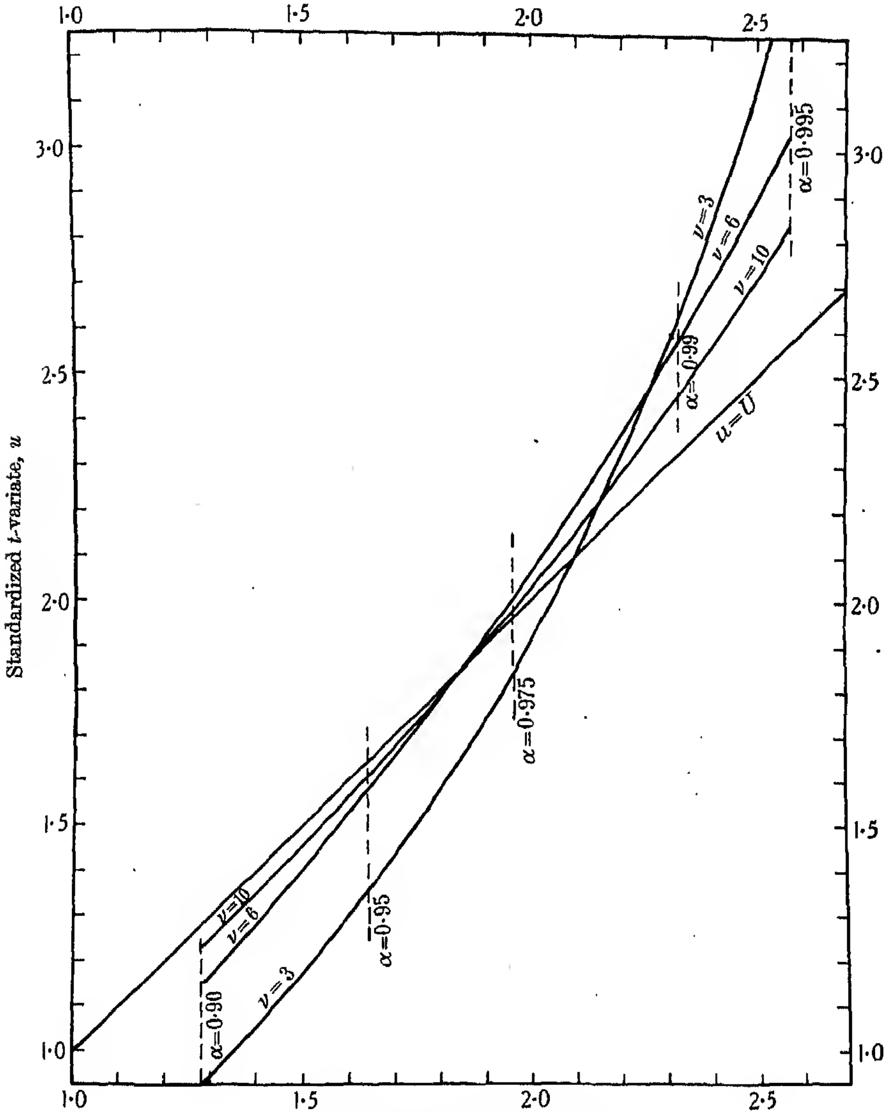


Fig. 3. Relation between $u(t)$ and U.

It remains to point out that, if the n tabulated probability levels are denoted by

$$\alpha_1 < \alpha_2 < \ldots < \alpha_n \leqslant 0{\cdot}10,$$

and if the interpolation is to be carried by using a quadratic or higher expression in the argument, it has been found that the interpolated value is always more

accurate when the probability levels used include as many as possible of the probability levels below α_x. Similarly, if

$$0{\cdot}90 \leqslant \alpha_1 < \alpha_2 < \ldots < \alpha_n,$$

it is better to use probability levels including as many as possible of those above α_x.

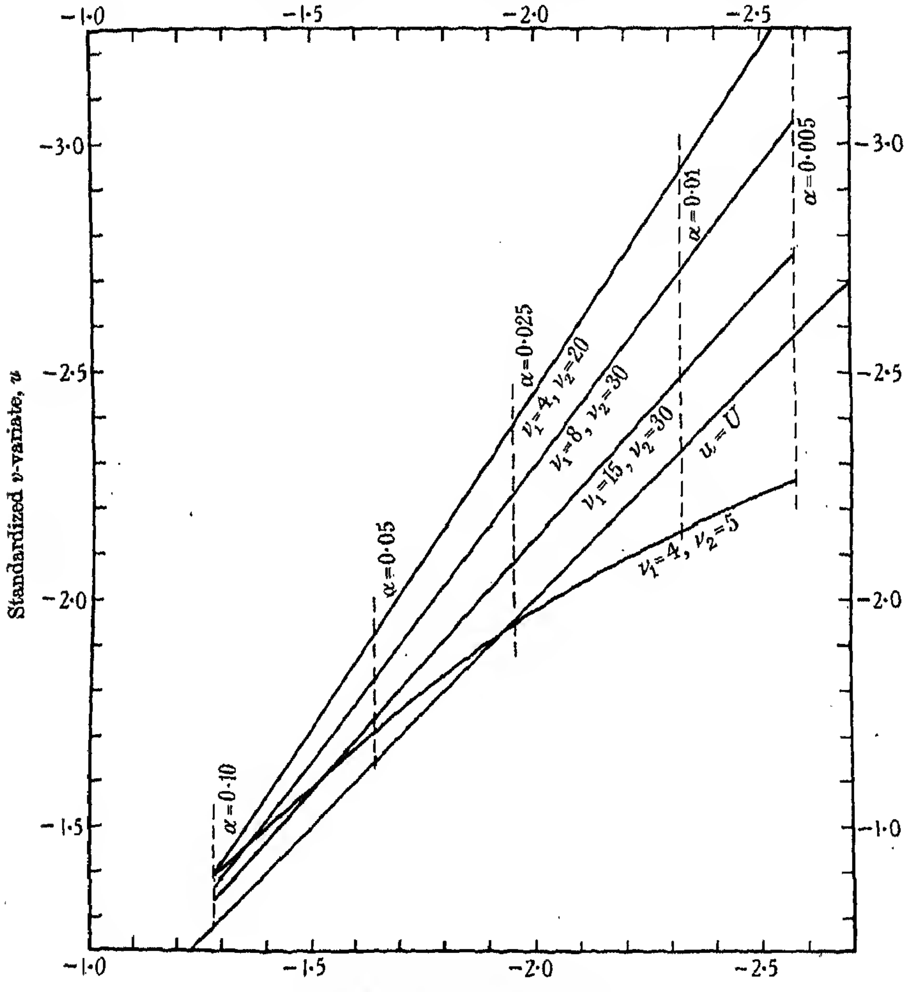


Fig. 4. Relation between $u(v)$ and U, where $p(v) = Cv^{\frac{1}{2}\nu_2-1}(1-v)^{\frac{1}{2}\nu_1-1}$.

4. The χ^2-probability function

The standardized form of χ^2 used in Fig. 1 is

$$u(\chi^2) = \frac{\chi^2 - \nu}{\sqrt{(2\nu)}}, \tag{6}$$

ν being the number of degrees of freedom. The γ_1 and γ_2 of this distribution are

$\sqrt{(8/\nu)}$ and $12/\nu$ respectively. If we consider χ itself instead of χ^2 we find that its standardized form has the approximate value, used in Fig. 2,

$$u(\chi) = \frac{\chi\sqrt{2} - \sqrt{(2\nu-1)}}{1 - \frac{1}{8}\nu^{-1}}, \tag{7}$$

and its γ_1 and γ_2 are $(2\nu)^{-\frac{1}{2}} + O(\nu^{-2})$ and $O(\nu^{-2})$ respectively. Both these last quantities are smaller than those for χ^2, which suggests that interpolation will be more accurate using χ rather than χ^2. This can also be seen from Figs. 1 and 2.

The χ^2 probability levels have been tabulated by Fisher (1941) for $\alpha = 0{\cdot}01$, $0{\cdot}02$, $0{\cdot}05$, $0{\cdot}10$, ..., $0{\cdot}90$, $0{\cdot}95$, $0{\cdot}98$, $0{\cdot}99$ and are given to three decimal places. These levels were used in the interpolation.* The accuracy of the interpolated values obtained by the eight methods detailed below was checked either from the *Tables of the Incomplete Gamma Function* (Karl Pearson, 1922) or from *Tables or Statisticians and Biometricians*, Part I, Table XII (Karl Pearson, 1930).

The greatest deviations, δ_m $(m = 1, 2, \ldots, 8)$, obtained using in all eight different methods, are given for $\nu = 3, 5, 9$ and 18 and for intervals of α: $(0{\cdot}01, 0{\cdot}02)$, $(0{\cdot}02, 0{\cdot}05)$, $(0{\cdot}05, 0{\cdot}10)$, $(0{\cdot}90, 0{\cdot}95)$, $(0{\cdot}95, 0{\cdot}98)$ and $(0{\cdot}98, 0{\cdot}99)$ in Table 1.

Method 1.

$$\chi^2 = A + B\log\alpha \quad (\alpha < 0{\cdot}10), \qquad \chi^2 = A + B\log(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 2.

$$u(\chi) = A + B\log\alpha \quad (\alpha < 0{\cdot}10), \qquad u(\chi) = A + B\log(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 3. $$u(\chi^2) = A + BU.$$

Method 4.

$$u(\chi^2) - U = A + B\log\alpha \quad (\alpha < 0{\cdot}10), \qquad u(\chi^2) - U = A + B\log(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 5. $$u(\chi) = A + BU.$$

Method 6. $$\chi^2 = A + B\log\alpha + C\log^2\alpha \quad (\alpha < 0{\cdot}10),$$
$$\chi^2 = A + B\log(1-\alpha) + C\log^2(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 7. $$u(\chi^2) = A + BU + CU^2.$$

Method 8. $$u(\chi) = A + BU + CU^2.$$

The best linear interpolation is that provided by method 5 and the best quadratic one is given by method 8. Both these methods are those suggested by the general approach.

* Certain additional levels are given in a recently published table computed by Catherine M. Thompson (1941).

Table 1. *Greatest deviations in the interpolation of χ^2 and χ.*

ν	γ_1	γ_2	Interval	α	True χ^2	$\lvert\delta_1\rvert$	$\lvert\delta_2\rvert$	$\lvert\delta_3\rvert$	$\lvert\delta_4\rvert$	$\lvert\delta_5\rvert$	$\lvert\delta_6\rvert$	$\lvert\delta_7\rvert$	$\lvert\delta_8\rvert$
3	1·63	4·00	(0·02, 0·05)	0·02 997	0·245	0·006	0·007	0·011	0·012	0·004	0·003	0·003	0·001
			(0·05, 0·10)	0·07 890	0·490	15	8	10	1	4	4	3	0
			(0·90, 0·95)	0·92 810	7·000	4	3	27	14	5	1	0	1
			(0·95, 0·98)	0·97 071	9·000	5	33	36	10	6	0	0	2
			(0·98, 0·99)	0·98 827	11·000	1	13	34	5	1	2	1	2
5	1·26	2·40	(0·01, 0·02)	0·01 353	0·632	8	5	6	3	2	6	0	0
			(0·02, 0·05)	0·03 340	0·949	23	13	15	3	5	5	1	0
			(0·05, 0·10)	0·06 150	1·265	19	11	11	0	3	4	0	1
			(0·90, 0·95)	0·92 808	10·119	11	32	26	12	5	3	1	1
			(0·95, 0·98)	0·96 544	12·017	13	39	32	12	6	2	1	0
			(0·98, 0·99)	0·98 579	14·230	5	18	17	6	4	2	1	1
9	0·94	1·33	(0·01, 0·02)	0·01 060	2·121	4	2	2	2	0	1	1	0
			(0·02, 0·05)	0·03 452	2·970	34	21	19	0	6	6	0	1
			(0·05, 0·10)	0·07 705	3·818	33	24	17	6	5	8	1	0
			(0·90, 0·95)	0·93 312	16·000	19	38	17	11	5	4	0	0
			(0·95, 0·98)	0·96 483	18·000	22	46	31	12	7	4	1	1
			(0·98, 0·99)	0·98 735	21·000	8	18	14	5	4	2	1	0
18	0·67	0·67	(0·01, 0·02)	0·01 167	7·200	14	9	5	2	1	4	1	1
			(0·02, 0·05)	0·02 793	8·400	47	32	20	2	5	7	0	1
			(0·05, 0·10)	0·07 482	10·200	48	34	18	5	6	11	1	0
			(0·90, 0·95)	0·93 159	27·600	34	52	25	10	5	7	1	1
			(0·95, 0·98)	0·96 255	30·000	34	55	28	10	6	6	1	1
			(0·98, 0·99)	0·98 589	33·600	15	26	13	5	4	3	1	1

For methods associated with subscripts to δ, see p. 269.

Table 2. *Greatest deviations in the interpolation of t*

ν	γ_2	Interval	α	True t	$\lvert\delta_1\rvert$	$\lvert\delta_2\rvert$	$\lvert\delta_3\rvert$	$\lvert\delta_4\rvert$	$\lvert\delta_5\rvert$	$\lvert\delta_6\rvert$	$\lvert\delta_7\rvert$	$\lvert\delta_8\rvert$
3	∞	(0·005, 0·01)	0·00 692	5·196	0·039	0·040	0·035	0·020	0·011	0·008	0·007	0·004
		(0·01, 0·025)	0·01 430	3·969	64	52	42	21	15	7	6	2
		(0·025, 0·05)	0·03 261	2·848	28	24	16	7	8	5	3	0
		(0·050, 0·10)	0·07 285	1·954	25	23	11	4	7	3	3	1
6	3·00	(0·005, 0·01)	0·00 714	3·413	11	8	5	3	2	1	1	1
		(0·01, 0·025)	0·01 517	2·820	18	14	6	3	2	1	1	1
		(0·025, 0·05)	0·03 874	2·128	8	6	1	1	2	2	0	1
		(0·050, 0·10)	0·06 820	1·719	9	8	1	1	1	1	0	0
10	1·00	(0·005, 0·01)	0·00 625	3·038	5	4	1	2	1	1	1	1
		(0·01, 0·025)	0·01 639	2·476	8	11	1	1	1	1	1	1
		(0·025, 0·05)	0·03 844	1·972	4	3	2	1	1	1	0	0
		(0·050, 0·10)	0·07 246	1·581	4	4	5	0	1	1	0	0

For methods associated with subscripts to δ, see p. 271.

5. THE t-PROBABILITY FUNCTION

The standardized form of the t-probability function used in Fig. 3 is

$$u(t) = t\sqrt{\left(\frac{\nu-2}{\nu}\right)}, \tag{8}$$

ν being the number of degrees of freedom. The values of γ_1 and γ_2 are zero and $6/(\nu-4)$ respectively.

Percentage levels for t have been tabulated by Fisher (1941) for $\alpha = 0{\cdot}005$, $0{\cdot}01$, $0{\cdot}025$, $0{\cdot}05$, $0{\cdot}10$, ... and are given to three decimal places. The level here defined as α is half the figure given in Fisher's table, i.e.

$$\alpha = \int_{-\infty}^{t_\alpha} p(t)\,dt. \tag{9}$$

The values of t used in checking those obtained by interpolation are taken from *Tables of the Incomplete Beta-Function* (Karl Pearson, 1934), where

$$\left.\begin{aligned} &I_x(p,q) = \alpha, \\ &p = \tfrac{1}{2}\nu, \quad q = \tfrac{1}{2} \quad \text{and} \quad x = \frac{1}{1+t^2/\nu} \end{aligned}\right\}, \tag{10}$$

ν being the number of degrees of freedom.

The greatest deviations found in the following eight methods are denoted by δ_m ($m = 1, 2, \ldots, 8$), and are given in Table 2 for $\nu = 3, 6$ and 10 and for the intervals $(0{\cdot}005, 0{\cdot}01)$, $(0{\cdot}01, 0{\cdot}025)$, $(0{\cdot}025, 0{\cdot}05)$, $(0{\cdot}05, 0{\cdot}10)$.

Method 1. $\quad u(t) = A + BU.$

Method 2. $\quad u(t) - U = A + B\log\alpha.$

Method 3. $\quad t = A + B\log\alpha.$

Method 4. $\quad u(t) = AU + BU^3.$

Method 5. $\quad u(t) = A + BU + CU^2.$

Method 6. $\quad t = A + B\log\alpha + C\log^2\alpha.$

Method 7. $\quad u(t) - U = A + B\log\alpha + C\log^2\alpha.$

Method 8. $\quad u(t) = AU + BU^3 + CU^5.$

From the methods that make use of only two probability levels, method 4 is the one that gives the highest accuracy and from those that make use of three probability levels, method 8 is the best. These two methods are those suggested by the general approach.

6. THE v OR BETA-DISTRIBUTION

This variable, which is related to R. A. Fisher's z, is defined as

$$v = \frac{S_2}{S_1+S_2} = \frac{\nu_2}{\nu_2+\nu_1 e^{2z}}, \tag{11}$$

where S_1 is a sum of squares of normal variates based on ν_1 degrees of freedom and S_2 an independent sum of squares based on ν_2 degrees of freedom. The elementary probability law of v is

$$p(v) = \{B(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2)\}^{-1} v^{\frac{1}{2}\nu_2-1}(1-v)^{\frac{1}{2}\nu_1-1}. \tag{12}$$

Table 3. *Greatest deviations in the interpolation of v*

ν_1	ν_2	γ_1	γ_2	Interval	α	True v	$\|\delta_1\|$	$\|\delta_2\|$	$\|\delta_3\|$	$\|\delta_4\|$	$\|\delta_5\|$	$\|\delta_6\|$	$\|\delta_7\|$
4	20	−0·92	+0·79	(0·005, 0·01)	0·00 838	0·52000	0·00043	0·00016	0·00006	0·00007	0·00006	0·00005	0·00008
				(0·01, 0·05)	0·02 607	0·59000	296	76	34	35	·7	27	1
				(0·05, 0·10)	0·06 901	0·66000	79	5	28	14	6	10	1
8	30	−0·61	+0·25	(0·005, 0·01)	0·00 769	0·53000	25	4	7	16	9	5	8
				(0·01, 0·05)	0·02 711	0·59000	228	36	24	35	4	13	1
				(0·05, 0·10)	0·06 645	0·64000	67	4	16	16	2	5	2
15	30	−0·28	−0·12	(0·005, 0·01)	0·00 666	0·41000	31	3	3	8	9	7	8
				(0·01, 0·05)	0·02 826	0·47000	240	40	22	42	3	9	3
				(0·05, 0·10)	0·07 463	0·52000	78	9	1	20	1	9	1
8	15	−0·33	−0·26	(0·005, 0·01)	0·00677	0·30000	53	19	20	13	3	3	3
				(0·01, 0·05)	0·02 519	0·37000	418	128	114	67	11	22	2
				(0·05, 0·10)	0·06 972	0·44000	126	27	16	29	2	15	3
4	5	−0·16	−0·77	(0·005, 0·01)	0·00 796	0·09000	80	32	58	29	16	5	5
				(0·01, 0·05)	0·02 723	0·15000	838	375	534	181	93	19	12
				(0·05, 0·10)	0·07 421	0·23000	278	123	144	78	33	8	11

For methods associated with subscripts to δ, see p. 273.

Low values of v correspond to high values of z. The standardized form of v used in Fig. 4 is

$$u(v) = \frac{v - \dfrac{\nu_2}{\nu_1+\nu_2}}{\dfrac{1}{\nu_1+\nu_2}\sqrt{\left(\dfrac{2\nu_1\nu_2}{\nu_1+\nu_2+2}\right)}}, \tag{13}$$

and the values of γ_1 and γ_2 are

$$\left.\begin{aligned} \gamma_1 &= \frac{\nu_1-\nu_2}{\nu_1+\nu_2+4}\sqrt{\left(\frac{8(\nu_1+\nu_2+2)}{\nu_1\nu_2}\right)}, \\ \gamma_2 &= \frac{12\{\nu_1^2(\nu_1+2)+\nu_2^2(\nu_2+2)-2\nu_1\nu_2(\nu_1+\nu_2+4)\}}{\nu_1\nu_2(\nu_1+\nu_2+4)(\nu_1+\nu_2+6)}. \end{aligned}\right\} \tag{14}$$

The cases considered here are those most generally met in tests of significance with $\nu_1 < \nu_2$ and therefore $\gamma_1 < 0$. γ_2 is sometimes positive and sometimes negative.

Tables giving values of v for $\alpha = 0{\cdot}005$, 0·01, 0·025, 0·05, 0·10, 0·25 and 0·50 to five decimal places have recently been published (Catherine M. Thompson, 1941). The corresponding upper probability levels can be obtained by entering the tables with ν_1 and ν_2 transposed and taking $1-v$ for v.

The values of v used in checking the interpolation are taken from the *Tables of the Incomplete Beta-Function* (Karl Pearson, 1934).

The greatest deviations δ_m $(m = 1, 2, \ldots, 7)$, found in the following seven methods are given in Table 3 for the following pairs of values of ν_1 and ν_2: (4, 20) (8, 30) (15, 30) (8, 15) (4, 5) and for the intervals (0·005, 0·01), (0·01, 0·05), (0·05, 0·10) of α.

Method 1.

$$v = A + B\log\alpha \quad (\alpha < 0{\cdot}10), \qquad v = A + B\log(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 2.

$$u(v) - U = A + B\log\alpha \quad (\alpha < 0{\cdot}10), \qquad u(v) - U = A + B\log(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 3.

$$u(v) = A + BU.$$

Method 4.

$$v = A + B\log\alpha + C\log^2\alpha \quad (\alpha < 0{\cdot}10),$$
$$v = A + B\log(1-\alpha) + C\log^2(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 5.

$$u(v) - U = A + B\log\alpha + C\log^2\alpha \quad (\alpha < 0{\cdot}10),$$
$$u(v) - U = A + B\log(1-\alpha) + C\log^2(1-\alpha) \quad (\alpha > 0{\cdot}90).$$

Method 6.

$$u(v) = A + BU + CU^2.$$

Method 7.

$$u(v) = A + BU + CU^2 + DU^3.$$

Here also the best linear method and the best quadratic one are those suggested by the general approach, namely, methods 3 and 6. Method 7, a cubic in U, was used only because the great accuracy of the tabulated v-function justified the computation. This cubic interpolation gives errors of the order of 0·0001 even for numbers of degrees of freedom as small as $\nu_1 = 4$, $\nu_2 = 5$. If the interval between $\alpha = 0{\cdot}01$ and 0·05 were broken into two parts at 0·025 (a level given in the Thompson tables) the corresponding errors δ_3 would be considerably reduced.

7. Some numerical examples, using linear interpolation

(*a*) *Interpolation for a* χ^2 *level.* Suppose that we calculate the upper $2\frac{1}{2}$ % level ($\alpha = 0{\cdot}975$) for χ^2 with $\nu = 5$ degrees of freedom, using tabled values of the upper 2 % and 5 % levels. We require the 2 %, $2\frac{1}{2}$ %, and 5 % levels of the standardized normal variable as well as the two χ^2 levels. The relevant data are shown in the following table; as pointed out above, there is no need to calculate the standardized form of either χ or χ^2.

α	U_α	χ^2_α	χ_α
0·98	2·0537	13·388	3·659
0·975	1·9600	?	?
0·95	1·6449	11·070	3·327

By linear interpolation, i.e. using method 3 of p. 269, we have

$$\chi^2_{0\cdot975} = 11\cdot070 + (13\cdot388 - 11\cdot070) \times \frac{1\cdot9600 - 1\cdot6449}{2\cdot0537 - 1\cdot6449}$$
$$= 12\cdot857.$$

Interpolating similarly for χ, i.e. using method 5 of p. 269, we find

$$\chi_{0\cdot975} = 3\cdot583 \quad \text{or} \quad \chi^2_{0\cdot975} = 12\cdot838.$$

The correct value taken from Miss Thompson's table (1941) is

$$\chi^2_{0\cdot975} = 12\cdot8325.$$

It is seen, as expected on theoretical grounds and as evidenced in the comparisons of Table 1, that method 5 is the more accurate of the two.

(*b*) *Interpolation for a t level.* Suppose that we calculate the value of t corresponding to $\alpha = 0\cdot0125$ (as defined in equation (9)), with $\nu = 6$, using the tabled levels for $\alpha = 0\cdot005$ and $0\cdot025$. The data required are shown below, the values of t being taken from the table on p. 300 of this issue.

α	U_α	t_α	t_α/U_α	U_α
0·005	−2·5758	−3·7074	1·4393	6·6347
0·0125	−2·2414	?	?	5·0239
0·025	−1·9600	−2·4469	1·2484	3·8416

Again, it is not necessary to calculate the standardized values of t_α, for even when using the relation of method 4, which assumes

$$\frac{t}{\sigma_t U} = A + BU^2, \tag{15}$$

the transference of σ_t to the right-hand side of the equation will only modify the constants A and B whose values are not directly determined in the interpolation process.

Using method 1 (t a linear function of U), it is found that

$$t_{0\cdot0125} = -3\cdot023.$$

Using method 4 (t/U a linear function of U^2), it is found that

$$t_{0\cdot0125} = -2\cdot979.$$

The correct value is $-2\cdot969$. As can be seen in Fig. 3, from the stretch of the curve for $\nu = 6$ between $\alpha = 0\cdot975$ and $0\cdot995$, the interval chosen is too long for satisfactory linear interpolation. The use of formula (15) improves matters, but is still not satisfactory. With Fisher's tables we could, of course, interpolate between the levels $\alpha = 0\cdot010$ and $0\cdot025$; doing this with method 4, it is found that

$$t_{0\cdot0125} = -2\cdot971$$

a distinctly better value.

(c) *Interpolation for a beta-distribution percentage level.* Take the case $\nu_1 = 8$, $\nu_2 = 30$ and suppose it is wished to find the $2\frac{1}{2}$ % level from a knowledge of the 1 % and 5 % levels. The data required are as follows, $v_{0\cdot 01}$ and $v_{0\cdot 05}$ being taken from Miss Thompson's tables.

α	U_α	v_α
0·01	2·3263	0·54170
0·025	1·9600	?
0·05	1·6449	0·62332

Then, using linear interpolation, i.e. method 3 of p. 273, we have

$$u_{0\cdot 025} = 0\cdot 62332 + (0\cdot 62332 - 0\cdot 54170) \times \frac{1\cdot 9600 - 1\cdot 6449}{2\cdot 3263 - 1\cdot 6449}$$
$$= 0\cdot 58558.$$

The correct value taken from the same tables is 0·58582. It will be seen from Fig. 4 that the intervals $\alpha = 0\cdot 005, 0\cdot 010, 0\cdot 025, 0\cdot 050, 0\cdot 100$ of Miss Thompson's tables are likely to form a satisfactory framework for subtabulation if this is needed.

8. SUBTABULATION OF EXISTING TABLES

These methods have been used to produce the following enlargements of existing tables; but it is not at present possible to arrange for their publication.

(a) *Table of χ^2 percentage levels.* Method 7 (p. 269) was used, as it is almost as accurate as method 8 and less laborious. The table calculated gives χ^2_α to 3 decimal places for $\nu = 1\,(1)\,30$, and for $\alpha = 0\cdot 010\,(0\cdot 002)\,0\cdot 020;\ 0\cdot 020\,(0\cdot 005)\,0\cdot 050;\ 0\cdot 05\,(0\cdot 01)\,0\cdot 10$; and for the corresponding levels at the upper end of the distribution, i.e. for $\alpha' = 1 - \alpha$.

(b) *Table of t percentage levels.* Method 8 (p. 271) was used for $\nu = 3, \ldots, 6$, and method 6 for $\nu = 7\,(1)\,30$; 40, 60 and 120. Exact levels were calculated for $\nu = 1, 2$. The tables were computed to 3 decimal places and for

$$\alpha = 0\cdot 005\,(0\cdot 001)\,0\cdot 010;\ 0\cdot 0100\,(0\cdot 0025)\,0\cdot 0250;\ 0\cdot 025\,(0\cdot 005)\,0\cdot 050;\ 0\cdot 05\,(0\cdot 01)\,0\cdot 10.$$

9. INTERPOLATION FOR THE PROBABILITY INTEGRAL α, GIVEN u_α.

As we have already mentioned, the variable U_α can be expressed as a polynomial in u_α. This suggests that interpolation for U_α, given u_α, is as easy and as accurate as the interpolation discussed above. Hence to interpolate for α, given the value of u_α and certain tabled levels $u_{\alpha_1}, u_{\alpha_2}, \ldots$, we first interpolate for U_α and then find the value of α from appropriate tables of the normal probability integral, e.g. *Tables for Statisticians and Biometricians*, Part I, Table II.

10. Conclusion

It has been shown how accurate values of the probability levels of a statistical variate may be interpolated between standard tabled values by using the standardized normal variate as auxiliary. For many purposes linear interpolation is adequate; for others a second order Lagrangian formula may be preferred. The accuracy of the result depends, of course, on the closeness of the actual probability law to the normal law and on the size of the intervals between the tabled levels. The method has been illustrated on examples from the χ^2, t and v (beta) distributions.

The method has been used to provide a subtabulation of existing tables of percentage points for χ^2 and t, but it is not possible to have these tables printed with the present contribution.

Finally I should like to express my thanks to Dr B. L. Welch of University College, London, for originally suggesting the problem to me.

References

Cornish, E. A. & Fisher, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* no. 4.

Fisher, R. A. (1941). *Statistical Methods for Research Workers*, 8th Edition. Edinburgh: Oliver & Boyd.

Pearson, K. (1922). *Tables of the Incomplete Gamma-Function.*

—— (1930). *Tables for Statisticians and Biometricians*, Part I.

—— (1934). *Tables of the Complete and Incomplete Beta-Function.*

Thompson, Catherine M. (1941). *Biometrika*, **32**, 168–81

PARTIAL RANK CORRELATION

By M. G. KENDALL

1. In interpreting an observed dependence between two qualities we are constantly faced with the question whether an association (correlation) of A with B is really due to the associations (correlations) of each with a third quality C. This has led naturally to the theories of partial association and correlation, which attempt to decide the matter by the consideration of subpopulations in which the variation of C is eliminated. An analogous problem arises in ranking work but, so far as I know, has not previously been considered. For example, if a number of men are ranked according to mathematical and musical aptitude and there appears a significant rank correlation, it is natural to inquire whether this may be attributable to the correlation of both with some more fundamental quality such as intelligence. The object of this paper is to propose a coefficient of partial rank correlation which has a natural meaning and may be found useful for investigations requiring this type of decision.

2. As a preliminary it may be worth examining what can be done in this direction with the Spearman rank correlation coefficient ρ. If there are three rankings denoted by 1, 2 and 3, we may find the three coefficients ρ_{12}, ρ_{13}, ρ_{23}. It is tempting to apply to these coefficients the formulae of product moment partial correlation such as

$$\rho_{23.1} = \frac{\rho_{23}-\rho_{12}\rho_{13}}{(1-\rho_{12}^2)^{\frac{1}{2}}(1-\rho_{13}^2)^{\frac{1}{2}}}, \tag{1}$$

and to *define* $\rho_{23.1}$ as the partial rank correlation of 2 and 3 'when 1 is constant'. There is clearly very little justification for such a procedure, and it is far from easy to explain just what $\rho_{23.1}$ means. In fact, the only defence of formula (1) that can bear critical examination is, I think, that it is an approximation to a second possibility, as follows:

3. There can be no such thing as a rank correlation in a continuous population (the members of which are not even denumerable) but we can speak with genuine meaning of a grade correlation. A well-known result due to Karl Pearson states that in a normal bivariate population with correlation ρ_p the grade correlation ρ_g is given by

$$\rho_p = 2\sin\frac{\pi}{6}\rho_g. \tag{2}$$

The Spearman coefficient ρ may be regarded as a sample grade correlation. If, therefore, we take ρ as an estimate of ρ_g we may find ρ_p from (2). For three rankings we may then obtain the three values of ρ_p, apply the ordinary product moment partial formula, and so obtain a partial coefficient. Since x and $2\sin\frac{\pi x}{6}$ do not

differ by more than a small amount for $|x| \leqslant 1$; we might even apply formula (1) direct to the values of ρ without bothering to transform them into ρ_p by equation (2).

4. Such a procedure, again, is open to fairly obvious objections. Apart from the all-too-facile assumption of normality and the error involved in using Spearman's ρ from a small sample to estimate the grade correlation in a parent, the fact remains that we arrive, not at a partial rank coefficient, but at an estimate of a partial product moment coefficient in a normal population.

Perhaps there are cases where this is a reasonable objective based on reasonable assumptions but it is evidently unsatisfactory for general ranking purposes.

5. In a previous paper (1938) I defined an alternative coefficient of rank correlation τ which may be generalized to include the case when pairs of individuals are compared separately (Kendall & Babington Smith, 1940). It will be convenient for present purposes to redefine τ in a slightly different manner so that the results obtained below may again be immediately generalized to the case of paired comparisons. Consider the two rankings of six

$$\begin{array}{lcccccc} 1: & 1 & 4 & 3 & 2 & 6 & 5 \\ 2: & 3 & 2 & 4 & 1 & 5 & 6. \end{array} \tag{3}$$

There are $\binom{6}{2} = 15$ possible pairs of ranks in each ranking, viz. 12, 13, ..., 16, 23, 24, ..., 56. We write them down as in expression (4) below. Any order of the pairs will serve, and it is immaterial whether any pair is written as ab or ba; but for practical convenience they may be written in the natural order indicated in the last sentence but one. This arrangement I call the recorded order.

We then consider the occurrence of each pair in the ranking 1. If a pair occurs in that ranking in the order in which we have recorded it, we write a plus below the recorded order underneath the pair concerned; in the contrary case we write a minus. Ranking 1 of expression (3) will then give

Recorded order:	(12)	(13)	(14)	(15)	(16)	(23)	(24)	(25)	(26)	(34)	(35)	(36)	(45)	(46)	(56)
Ranking 1:	+	+	+	+	+	−	−	+	+	−	+	+	+	+	−

(4)

Here, for example, the pair (15) occurs in that order in ranking 1 and so is denoted by a +, whereas the pair (24) occurs as 42 and is denoted by a −.

Consider now ranking 2. The members of ranking 1 which are ranked 1, 2 correspond to members in ranking 2 ranked as 3, 1. This is in the reverse order to that of the pair 13 in the recorded order, so (starting a new row of signs corresponding to ranking 2) we write a minus under the recorded pair (12). Similarly, the pair in ranking 2 corresponding to 15 in ranking 1 is 36. This is in the same order as the recorded pair, so we write a plus below the existing plus under the recorded pair (15). The pair in ranking 2 corresponding to 23 in ranking 1 is 41. This is in the reverse order of the recorded pair, so we write a minus below the

recorded pair (23) in the row of signs corresponding to ranking 2. And so on. This takes rather a long time to explain but the process is really very simple. The array corresponding to expressions (3) is then

Recorded order:	(12)	(13)	(14)	(15)	(16)	(23)	(24)	(25)	(26)	(34)	(35)	(36)	(45)	(46)	(56)
Ranking 1:	+	+	+	+	+	−	−	+	+	−	+	+	+	+	−
Ranking 2:	−	+	−	+	+	−	−	+	+	+	+	+	+	+	+

(5)

Now in expression (5) there are eleven cases in which both rankings have the same sign and therefore $15-11=4$ in which they have the opposite sign. The coefficient τ is then given by

$$\tau_{12}=\frac{11-4}{15}=\frac{7}{15}.$$

Generally, if there are, in two rankings of n arrayed as above, S_1 cases of the same sign and S_2 of opposite sign

$$\tau=\frac{2(S_1-S_2)}{n(n-1)}$$

$$=\frac{4S_1}{n(n-1)}-1=1-\frac{4S_2}{n(n-1)}. \tag{6}$$

If we arrange ranking 1 in the natural order $1, \ldots, n$, then every case in which there are the same signs in expression (5) corresponds to a case in which pairs in ranking 2 are in the natural order; and every case of different sign to one in which the pairs in ranking 2 are in the reverse of the natural order. The definition of (6) thus accords with the one originally given in my 1938 paper. It is often convenient to take one ranking to be the natural order $1, \ldots, n$ so that the first row of signs in (5) are all positive. For instance, on rearrangement of the rankings in (3) we have

1:	1	2	3	4	5	6
2:	3	1	4	2	6	5

(7)

and the array of paired comparisons becomes

Recorded order:	(12)	(13)	(14)	(15)	(16)	(23)	(24)	(25)	(26)	(34)	(35)	(36)	(45)	(46)	(56)
Ranking 1:	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ranking 2:	−	+	−	+	+	+	+	+	+	−	+	+	+	+	−

(8)

Here again $S_1=11$, $S_2=4$, $\tau_{12}=\frac{7}{15}$.

6. Consider now three rankings, of which the first may be taken to be the natural order $1, \ldots, n$, for example,

1:	1	2	3	4	5	6
2:	3	1	4	2	6	5
3:	4	2	1	6	3	5

(9)

The corresponding array of pairs is

Recorded order:	(12)	(13)	(14)	(15)	(16)	(23)	(24)	(25)	(26)	(34)	(35)	(36)	(45)	(46)	(56)
Ranking 1:	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ranking 2:	−	+	−	+	+	+	+	+	+	−	+	+	+	+	−
Ranking 3:	−	−	+	−	+	−	+	+	+	+	+	+	−	−	+

(10)

For the coefficients τ we have

$$\tau_{12} \text{ (as above)} = \frac{7}{15},$$

$$\tau_{13} = \frac{9-6}{15} = \frac{3}{15},$$

$$\tau_{23} = \frac{7-8}{15} = -\frac{1}{15},$$

in the last case S_1 being as usual the number of cases in which pairs in rankings 2 and 3 have the same sign.

Consider now the fourfold table setting out the occurrences of + and − signs in the rows of expression (10) corresponding to rankings 2 and 3:

Ranking 2

Ranking 3		+	−	Total
	+	6	5	11
	−	3	1	4
	Total	9	6	15

(11)

Here, for example, there are six cases in which pairs of ranks have the same (positive) sign, five in which ranking 2 is negative while ranking 3 is positive, and so on.

Generally, if for three rankings of n the table is of the form

Ranking 2

Ranking 3		+	−	Total
	+	a	b	$a+b$
	−	c	d	$c+d$
	Total	$a+c$	$b+d$	N

(12)

I define the partial rank correlation of 2 and 3 on 1 as

$$\tau_{23.1} = \frac{ad-bc}{+\sqrt{\{(a+b)(c+d)(a+c)(b+d)\}}} \tag{13}$$

$$= \sqrt{\frac{\chi^2}{N}}, \tag{14}$$

where $N = \binom{n}{2}$ and χ^2 is the ordinary mean square contingency for the fourfold table.

In the particular case here considered the coefficient is

$$\tau_{23.1} = \frac{6-15}{\sqrt{(4.11.9.6)}} = -0{\cdot}185,$$

as compared with $\tau_{23} = -0{\cdot}067.$

In the case when ranking 1 is not the natural order 1, ..., n the same principles apply, but in considering the communalities of 2 and 3 we count as contributing to the a-cell in (12) the pairs which are themselves of the same sign and are also of the same sign as the term in ranking 1; and so on.

7. The partial τ defined in equation (13) is a coefficient of association in the 2×2 table suggested by Yule (1912). When the attributes of the table are independent and only in this case it is zero. It is $+1$ only if b and c are both zero (i.e. if the two rankings agree in all pairs and hence are identical) and -1 only if a and d are both zero (i.e. if the two rankings disagree in all pairs and one is the reverse of the other). In this latter property of attaining unity only when diagonally opposite cells are both empty it differs from two other coefficients proposed by Yule.*

Thus partial τ as defined is a measure of the association of agreements of the rankings 2 and 3 when compared in pairs with ranking 1. From this viewpoint it will be seen that the use of the word 'partial' conforms to that in the theories of association and correlation. The partial associations of A and B in a third population containing C and γ are those of AC and BC or of $A\gamma$ and $B\gamma$. The partial association of ranks is that of pairs of agreements of 21 and 31. The partial correlation of 2 and 3 when 1 is constant is paralleled by the partial rank correlation of 2 and 3 when 1 is in the natural order 1, ..., n. If partial τ is unity in absolute value the rankings coincide or one is the reverse of the other, whether they agree with ranking 1 or not, so that the coefficient fulfils its proper function of measuring the relationship between 2 and 3 independently of the influence of 1.

* Yule himself arrived at the coefficient by considering product-moments in a 2×2 table. Karl Pearson & Heron (1913) mistook Yule's intention and thought the coefficient was proposed as an estimate of the correlation in a normal population whose frequencies were given by a double dichotomy in the 2×2 table. Their long memoir is mainly devoted to advocating the alternative merits of tetrachoric r. The two things, as Yule has emphasized, are quite different. I mention the point to make it clear that the Pearson-Heron criticisms of Yule's coefficient, even if not misfounded, do not affect the above work, since I use the coefficient purely as a measure of association in the fourfold table.

8. We may establish the remarkable result

$$\tau_{23.1} = \frac{\tau_{23} - \tau_{12}\tau_{13}}{(1-\tau_{12}^2)^{\frac{1}{2}}(1-\tau_{13}^2)^{\frac{1}{2}}}. \tag{15}$$

In fact, from expression (12) we see that

$$\tau_{12} = \frac{(a+b)-(c+d)}{N},$$

$$\tau_{13} = \frac{(a+c)-(b+d)}{N},$$

$$\tau_{23} = \frac{(a+d)-(b+c)}{N}.$$

Remembering that $N = a+b+c+d$ we have

$$1-\tau_{12}^2 = \frac{4}{N^2}(a+b)(c+d),$$

$$1-\tau_{13}^2 = \frac{4}{N^2}(a+c)(b+d),$$

$$\begin{aligned}\tau_{23}-\tau_{12}\tau_{13} &= \frac{1}{N^2}[(a+b+c+d)\{(a+d)-(b+c)\} \\ &\qquad -\{(a+b)-(c+d)\}\{(a+c)-(b+d)\}] \\ &= \frac{4}{N^2}(ad-bc).\end{aligned}$$

Equation (15) follows at once from the definition of partial τ in equation (13).

The appearance of the product-moment type of relation between total and partial correlations is surprising. There was no reason to expect that partial τ, which is a pure function of disarrangements in rankings and is not expressible algebraically in terms of the ranks, should bear any analogy with the partial correlation of variates; but since it does so, we are evidently fortified in regarding partial τ as a convenient measure of rank correlation.

Example. Ten men are ranked according to (1) intelligence, (2) mathematical ability, (3) musical ability. The rankings are:

1:	1	2	3	4	5	6	7	8	9	10
2:	1	4	5	6	2	7	3	9	8	10
3:	4	1	3	5	2	6	7	10	9	8

It will be found that $\tau_{12} = 0{\cdot}644$, $\tau_{13} = 0{\cdot}644$, $\tau_{23} = 0{\cdot}555$. Thus mathematical and musical ability are positively correlated. The question is, can this correlation be attributed to the correlation of both with intelligence?

We find
$$\tau_{23.1} = \frac{0{\cdot}555-(0{\cdot}644)^2}{1-(0{\cdot}644)^2} = 0{\cdot}24.$$

The conclusion would be that although part of the total correlation is due to the correlation of both with intelligence, part of it is not. We cannot attribute the whole of the observed (total) correlation between mathematical and musical ability to the interference of common correlation with intelligence.

9. The same methods are immediately capable of extension to paired comparisons. In fact the array of type (10) is the array of paired comparisons for all the possible pairs of ranks and when there are no constraints of the ranking character (i.e. such that if $A \to B$ and $B \to C$ then must $A \to C$) the coefficient τ becomes a measure of agreement in paired comparisons (cf. Kendall & Babington Smith, 1940). We could then construct measures of partial agreement by the same formulae in cases where it was suspected that there were mutual influences at work between three observers, as for instance if it were suspected that community of preference between two children was due to community of both with one of the parents.

10. In conclusion, it may perhaps be desirable to point out that although partial τ is defined by equation (14) in terms of χ^2 for a fourfold table, its significance cannot on that account be tested in the Type III χ^2 distribution with one degree of freedom. I have not yet succeeded in finding expressions for the sampling distribution of partial τ but it seems clear that the Type III distribution will not be reproduced in the ranking case, at least without some substantial modification, even when the rankings are independent; for there exist correlations between the signs given by any ranking in the recorded order. If, for example, (12) and (23) are positive so must be (13), whereas if (13) and (23) are positive (12) may be either positive or negative. The units in the fourfold table cannot therefore be regarded as allocated at random and the type III distribution will probably not hold. I hope to return to this subject on a later occasion.

REFERENCES

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

KENDALL, M. G. & BABINGTON SMITH, B. (1940). On the method of paired comparisons. *Biometrika*, **31**, 324.

PEARSON, KARL & HERON, D. (1913). On theories of association. *Biometrika*, **9**, 159.

YULE, G. UDNY (1912). On the methods of measuring the association between two attributes. *J.R. Statist. Soc.* **75**, 579.

INEQUALITIES FOR MULTIVARIATE FREQUENCY DISTRIBUTIONS

By C. E. V. LESER

GIVEN a frequency distribution y of a single variate x, with arithmetic mean $\bar{x} = 0$ and with standard deviation σ, Tchebycheff's (1867) well-known inequality presents a lower limit for the ratio of the frequency of all values of x between $-\lambda\sigma$ and $\lambda\sigma$ to the total frequency, where $\lambda \geqq 1$. In the case of the special class of frequency distributions for which $y(x)+y(-x)$ is a non-increasing function of $|x|$, this inequality can be substantially improved to another limit which applies to all positive values of λ, as already Gauss (1823) has proved. Various authors* have generalized these theorems by modifying the assumptions made with regard to the frequency distribution, by introducing moments of higher order than the second or by extending some of the results to bivariate functions. In the present analysis only moments of second order are used, but the results apply to frequency functions of any number of variates.

Suppose $y(x_1, \ldots, x_n)$ to be a frequency distribution of n variates, with total frequency equal to unity, arithmetic mean at the origin and with standard deviations $\sigma_1, \ldots, \sigma_n$. Let P be the frequency of all combinations of $x_1, \ldots, x_n$ for which

$$\left(\frac{x_1}{\lambda_1\sigma_1}\right)^2 + \ldots + \left(\frac{x_n}{\lambda_n\sigma_n}\right)^2 \leqq n.$$

Write
$$\lambda_0 = \sqrt{\frac{n}{1/\lambda_1^2 + \ldots + 1/\lambda_n^2}}, \quad \sigma_0 = \sqrt{\frac{n}{1/\sigma_1^2 + \ldots + 1/\sigma_n^2}},$$

so that λ_0^2 and σ_0^2 are the harmonic averages of $\lambda_1^2, \ldots, \lambda_n^2$ and $\sigma_1^2, \ldots, \sigma_n^2$ respectively. We also write

$$\mu_i = \frac{\lambda_i}{\lambda_0}, \quad \tau_i = \frac{\sigma_i}{\sigma_0} \quad (i = 1, \ldots, n),$$

$$R = \sqrt{\left\{\left(\frac{x_1}{\mu_1\tau_1}\right)^2 + \ldots + \left(\frac{x_n}{\mu_n\tau_n}\right)^2\right\}} = \lambda_0\sigma_0\sqrt{\left\{\left(\frac{x_1}{\lambda_1\sigma_1}\right)^2 + \ldots + \left(\frac{x_n}{\lambda_n\sigma_n}\right)^2\right\}},$$

so that P is the frequency of all values of $x_1, \ldots, x_n$ for which $R/\sqrt{n} \leqq \lambda_0\sigma_0$. Furthermore, we define $A(R_0)$ as the average value of y for all those values of $x_1, \ldots, x_n$ for which R has a fixed value R_0, and therefore

$$A(R_0) = \frac{\int_{R=R_0} \ldots \int y\, dx_1 \ldots dx_n}{\int_{R=R_0} \ldots \int dx_1 \ldots dx_n}.$$

* E.g. K. Pearson (1919), B. H. Camp (1922), S. Narumi (1923), C. D. Smith (1930), P. O. Berge (1937).

We can now state the following:

Theorem. Assuming the frequency distribution to be such that $A(R)$ is a non-increasing function of R for $R/\sqrt{n} \leqq \kappa\sigma_0$, we have one of the following three sets of inequalities, according to the value of κ:

I. $\kappa \leqq 1$:

(i) $\lambda_0 \leqq 1 \qquad P \geqq 0,$

(ii) $1 \leqq \lambda_0 \qquad P \geqq 1 - \dfrac{1}{\lambda_0^2}.$

II. $1 \leqq \kappa \leqq \left(\dfrac{n+2}{n}\right)^{\frac{1}{2}}$:

(i) $\lambda_0 \leqq \left(\dfrac{2}{n+2}\right)^{1/n} \kappa \qquad P \geqq \dfrac{n+2}{2}\left(1 - \dfrac{1}{\kappa^2}\right)\left(\dfrac{\lambda_0}{\kappa}\right)^n,$

(ii) $\left(\dfrac{2}{n+2}\right)^{1/n} \kappa \leqq \lambda_0 \leqq \kappa \qquad P \geqq 1 - \dfrac{1}{\kappa^2},$

(iii) $\kappa \leqq \lambda_0 \qquad P \geqq 1 - \dfrac{1}{\lambda_0^2}.$

III. $\left(\dfrac{n+2}{n}\right)^{\frac{1}{2}} \leqq \kappa$:

(i) $\lambda_0 \leqq \left(\dfrac{2}{n+2}\right)^{1/n}\left(\dfrac{n+2}{n}\right)^{\frac{1}{2}} \qquad P \geqq \left(\dfrac{n}{n+2}\right)^{\frac{1}{2}n} \lambda_0^n,$

(ii) $\left(\dfrac{2}{n+2}\right)^{1/n}\left(\dfrac{n+2}{n}\right)^{\frac{1}{2}} \leqq \lambda_0 \leqq \left(\dfrac{2}{n+2}\right)^{1/n} \kappa \qquad P \geqq 1 - \left(\dfrac{2}{n+2}\right)^{2/n} \dfrac{1}{\lambda_0^2},$

(iii) $\left(\dfrac{2}{n+2}\right)^{1/n} \kappa \leqq \lambda_0 \leqq \kappa \qquad P \geqq 1 - \dfrac{1}{\kappa^2},$

(iv) $\kappa \leqq \lambda_0 \qquad P \geqq 1 - \dfrac{1}{\lambda_0^2}.$

Proof. We introduce the new system of co-ordinates

$$\begin{aligned}
x_1 &= \mu_1 \tau_1 R \cos t_1 \cos t_2 \cos t_3 \ldots \cos t_{n-1}, \\
x_2 &= \mu_2 \tau_2 R \sin t_1 \cos t_2 \cos t_3 \ldots \cos t_{n-1}, \\
x_3 &= \mu_3 \tau_3 R \qquad \sin t_2 \cos t_3 \ldots \cos t_{n-1}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
x_n &= \mu_n \tau_n R \qquad\qquad\qquad \sin t_{n-1},
\end{aligned}$$

so that $\quad y(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = z(R, t_1, \ldots, t_{n-1})\, R^{n-1} dR\, dt_1 \ldots dt_{n-1},$

where $\quad z(R, t_1, \ldots, t_{n-1}) = y(R, t_1, \ldots, t_{n-1}) \cos t_2 \cos^2 t_3 \ldots \cos^{n-2} t_{n-1}.$

We also write $$s = \sigma_0\sqrt{n},$$

and we are going to use the following abbreviating symbol

$$\int_T f(R,t)\,dt = \int_{t_1=0}^{2\pi}\int_{t_2=-\frac{1}{2}\pi}^{\frac{1}{2}\pi}\cdots\int_{t_{n-1}=-\frac{1}{2}\pi}^{\frac{1}{2}\pi} f(R,t_1,\ldots,t_{n-1})\,dt_1\ldots dt_{n-1}.$$

Then $$\int_{R=0}^{\infty}\int_T R^{n-1}z(R,t)\,dR\,dt = 1,$$

$$\int_{R=0}^{\lambda_0 s}\int_T R^{n-1}z(R,t)\,dR\,dt = P,$$

and as $A(R) = \text{const.}\int_T z(R,t)\,dt$, we have the condition that $\int_T z(R,t)\,dt$ is a non-increasing function of R for $R \leqq \kappa s$. Furthermore, let us write

$$G = \int_T z(\lambda_0 s, t)\,dt,$$

and later on $$u = \frac{G}{n}s^n.$$

Now start from the equation

$$\left(\frac{\sigma_1}{\mu_1\tau_1}\right)^2+\ldots+\left(\frac{\sigma_n}{\mu_n\tau_n}\right)^2$$
$$= \int_{x_1=-\infty}^{\infty}\cdots\int_{x_n=-\infty}^{\infty}\left[\left(\frac{x_1}{\mu_1\tau_1}\right)^2+\ldots+\left(\frac{x_n}{\mu_n\tau_n}\right)^2\right]y(x_1,\ldots,x_n)\,dx_1,\ldots,dx_n$$

which can be written

$$(\lambda_0\sigma_0)^2\left(\frac{1}{\lambda_1^2}+\ldots+\frac{1}{\lambda_n^2}\right) = \int_{R=0}^{\infty}\int_T R^{n+1}z(R,t)\,dR\,dt$$

or $$s^2 = I_1 + I_2,$$

$$I_1 = \int_{R=0}^{\lambda_0 s}\int_T R^{n+1}z(R,t)\,dR\,dt, \quad I_2 = \int_{R=\lambda_0 s}^{\infty}\int_T R^{n+1}z(R,t)\,dR\,dt.$$

According to the value of κs, three cases have to be distinguished.

$$(a)\quad \kappa s \geqq \left[(\lambda_0 s)^n + \frac{n}{G}(1-P)\right]^{1/n} \quad \text{or} \quad P \geqq 1-(\kappa^n - \lambda_0^n)\,u.$$

For $R \leqq \lambda_0 s$ $$\int_T z(R,t)\,dt \geqq G.$$

Hence $$I_1 \geqq G\int_{R=0}^{\lambda_0 s} R^{n+1}\,dR,$$

and incidentally $$P \geqq G\int_{R=0}^{\lambda_0 s} R^{n-1}\,dR = \lambda_0^n u.$$

To obtain a limit for I_2, we make use of the equation

$$\int_{R=\lambda_0 s}^{\infty}\int_T R^{n-1}z(R,t)\,dR\,dt = 1-P = G\int_{R=\lambda_0 s}^{\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}} R^{n-1}dR$$

or
$$\int_{R=\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}}^{\infty}\int_T R^{n-1}z(R,t)\,dR\,dt$$
$$=\int_{R=\lambda_0 s}^{\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}} R^{n-1}\left[G-\int_T z(R,t)\,dt\right]dR.$$

As the integrands are nowhere negative and R nowhere smaller in the integral on the left than in the integral on the right, we can write

$$\int_{R=\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}}^{\infty}\int_T R^{n+1}z(R,t)\,dR\,dt$$
$$\geqq\int_{R=\lambda_0 s}^{\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}} R^{n+1}\left[G-\int_T z(R,t)\,dt\right]dR,$$
$$I_2\geqq G\int_{R=\lambda_0 s}^{\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}} R^{n+1}\,dR,$$

and therefore

$$s^2\geqq G\int_{R=0}^{\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}} R^{n+1}dR=\frac{[G(\lambda_0 s)^n+n(1-P)]^{(n+2)/n}}{(n+2)\,G^{2/n}},$$
$$n^{2/n}(n+2)\,u^2\geqq[n(\lambda_0^n u+1-P)]^{(n+2)/n},$$
$$P\geqq 1+\lambda_0^n u-\left(\frac{n+2}{n}\right)^{n/(n+2)}u^{2/(n+2)}.$$

$$(b)\quad \lambda_0 s\leqq\kappa s\leqq\left[(\lambda_0 s)^n+\frac{n}{G}(1-P)\right]^{1/n}\quad\text{or}\quad P\leqq 1-(\kappa^n-\lambda_0^n)\,u,\quad \lambda_0\leqq\kappa.$$

As before, $$I_1\geqq G\int_{R=0}^{\lambda_0 s}R^{n+1}\,dR\quad\text{and}\quad P\geqq\lambda_0^n u.$$

Furthermore, we have the equation

$$\int_{R=\lambda_0 s}^{\infty}\int_T R^{n-1}z(R,t)\,dR\,dt=1-P=G\int_{R=\lambda_0 s}^{\kappa s}R^{n-1}dR+\left[1-P-\frac{G}{n}(\kappa^n-\lambda_0^n)\,s^n\right],$$
$$\int_{R=\kappa s}^{\infty}\int_T R^{n-1}z(R,t)\,dR\,dt=\int_{R=\lambda_0 s}^{\kappa s}R^{n-1}\left[G-\int_T z(R,t)\,dt\right]dR$$
$$+\left[1-P-\frac{G}{n}(\kappa^n-\lambda_0^n)\,s^n\right],$$

and, as before, it follows that

$$\int_{R=\kappa s}^{\infty}\int_T R^{n+1} z(R,t)\,dR\,dt \geqq \int_{R=\lambda_0 s}^{\kappa s} R^{n+1}\left[G-\int_T z(R,t)\,dt\right]dR$$
$$+(\kappa s)^2\left[1-P-\frac{G}{n}(\kappa^n-\lambda_0^n)s^n\right],$$

$$I_2 \geqq G\int_{R=\lambda_0 s}^{\kappa s} R^{n+1}dR+(\kappa s)^2\left[1-P-\frac{G}{n}(\kappa^n-\lambda_0^n)s^n\right].$$

Hence
$$s^2 \geqq G\int_{R=0}^{\kappa s} R^{n+1}dR+(\kappa s)^2\left[1-P-\frac{G}{n}(\kappa^n-\lambda_0^n)s^n\right],$$

$$1 \geqq \frac{n}{n+2}\kappa^{n+2}u+\kappa^2[1-P-(\kappa^n-\lambda_0^n)u],$$

$$P \geqq \left(1-\frac{1}{\kappa^2}\right)+\left(\lambda_0^n-\frac{2}{n+2}\kappa^n\right)u.$$

(c)
$$\kappa s \leqq \lambda_0 s \quad \text{or} \quad \lambda_0 \geqq \kappa.$$

In this case, I_1 need not be larger than 0, but

$$I_2 \geqq (\lambda_0 s)^2(1-P),$$
$$s^2 \geqq (\lambda_0 s)^2(1-P).$$

Therefore, if $\lambda_0 \geqq 1$, $P \geqq 1-\frac{1}{\lambda_0^2}$, and if $\lambda_0 \leqq 1$, no significant limit exists.

For $\lambda_0 \geqq \kappa$, the problem has thus been solved, but the case $\lambda_0 \leqq \kappa$ is more complicated; summarizing the results of (*a*) and (*b*), we have two alternative sets of inequalities. Write

$$f_1(u)=\lambda_0^n u, \quad f_2(u)=1-(\kappa^n-\lambda_0^n)u,$$
$$f_3(u)=1+\lambda_0^n u-\left(\frac{n+2}{n}\right)^{n/(n+2)} u^{2/(n+2)},$$
$$f_4(u)=\left(1-\frac{1}{\kappa^2}\right)+\left(\lambda_0^n-\frac{2}{n+2}\kappa^n\right)u.$$

Then either
$$P \geqq f_1, \quad P \geqq f_2, \quad P \geqq f_3,$$
or
$$P \geqq f_1, \quad P \leqq f_2, \quad P \geqq f_4.$$

For different frequency distributions, G and therefore u may assume any non-negative value which is compatible with the condition $P \leqq 1$. We have to find the effective lower limit for P as a function of u and its minimum value which, as easily seen, is only larger than 0, if $f_4(0)>0$ or, what is the same thing, $\kappa>1$.

f_1, f_2, f_4 are straight lines. As

$$\frac{df_3}{du}=\lambda_0^n-\frac{2}{n+2}\left(\frac{n+2}{n}\right)^{n/(n+2)} u^{-n/(n+2)}$$

f_3 has a minimum when

$$u_3^{(\text{min.})} = \frac{n+2}{n}\left(\frac{2}{n+2}\right)^{(n+2)/n}\frac{1}{\lambda_0^{n+2}} = \frac{2}{n}\left(\frac{2}{n+2}\right)^{2/n}\frac{1}{\lambda_0^{n+2}}$$

and $$f_3^{(\text{min.})} = 1+\frac{2}{n}\left(\frac{2}{n+2}\right)^{2/n}\frac{1}{\lambda_0^2}-\frac{n+2}{n}\left(\frac{2}{n+2}\right)^{2/n}\frac{1}{\lambda_0^2} = 1-\left(\frac{2}{n+2}\right)^{2/n}\frac{1}{\lambda_0^2}.$$

Indicating by u_{ij}, f_{ij} the co-ordinates of the point in which the curves f_i and f_j intersect, we find that

$$u_{12} = \frac{1}{\kappa^n}, \qquad f_{12} = \left(\frac{\lambda_0}{\kappa}\right)^n,$$

$$u_{13} = \left(\frac{n}{n+2}\right)^{\frac{1}{2}n}, \qquad f_{13} = \left(\frac{n}{n+2}\right)^{\frac{1}{2}n}\lambda_0^n,$$

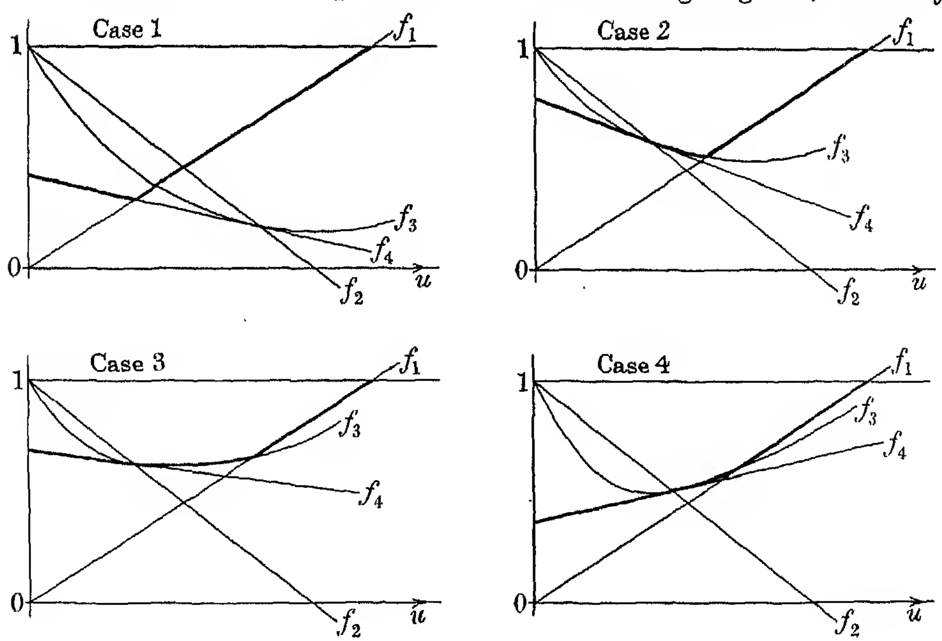
$$u_{14} = \frac{n+2}{2}\left(1-\frac{1}{\kappa^2}\right)\frac{1}{\kappa^n}, \qquad f_{14} = \frac{n+2}{2}\left(1-\frac{1}{\kappa^2}\right)\left(\frac{\lambda_0}{\kappa}\right)^n.$$

f_2, f_3 and f_4 have a point (u_{234}, f_{234}) in common in which f_4 is the tangent to f_3:

$$u_{234} = \frac{n+2}{n}\frac{1}{\kappa^{n+2}}, \qquad f_{234} = 1-\frac{n+2}{n}\frac{\kappa^n-\lambda_0^n}{\kappa^{n+2}}.$$

It is also seen that the sign of both expressions $\dfrac{df_4}{du}$ and $u_{234}-u_3^{(\text{min.})}$ is positive or negative according to whether $\lambda_0 \gtrless \left(\dfrac{2}{n+2}\right)^{1/n}\kappa$.

We have therefore four possibilities. In the following diagrams, the heavy



line shows the effective lower limit for P as a function of u (the upper limit is always equal to 1, of course), and its minimum gives the general lower limit for P.

In these four cases, the following results are obtained:

$$\begin{array}{llllll}
\left.\begin{array}{l}(1)\\(2)\\(3)\end{array}\right\} & \dfrac{df_4}{du} \leqq 0, & \left\{\begin{array}{l} u_{234} \geqq u_{12}, \\ u_{234} \leqq u_{12}, \end{array}\right. & \begin{array}{l} \\ u_3^{(\text{min.})} \geqq u_{13} \\ u_3^{(\text{min.})} \leqq u_{13} \end{array} & \begin{array}{l} P \geqq f_{14}, \\ P \geqq f_{13}, \\ P \geqq f_3^{(\text{min.})} \end{array} \\
(4) & \dfrac{df_4}{du} \geqq 0 & & & P \geqq f_4(0),
\end{array}$$

and by substituting the proper values:

$$\begin{array}{llll}
(1) & 1 \leqq \kappa \leqq \left(\dfrac{n+2}{n}\right)^{\frac{1}{2}}, & \lambda_0 \leqq \left(\dfrac{2}{n+2}\right)^{1/n} \kappa & P \geqq \dfrac{n+2}{2}\left(1-\dfrac{1}{\kappa^2}\right)\left(\dfrac{\lambda_0}{\kappa}\right)^n, \\
(2) & \multirow{2}{*}{$\kappa \geqq \left(\dfrac{n+2}{n}\right)^{\frac{1}{2}},$} & \lambda_0 \leqq \left(\dfrac{2}{n+2}\right)^{1/n}\left(\dfrac{n+2}{n}\right)^{\frac{1}{2}} & P \geqq \left(\dfrac{n}{n+2}\right)^{\frac{1}{2}n} \lambda_0^n \\
(3) & & \left(\dfrac{2}{n+2}\right)^{1/n} \kappa \geqq \lambda_0 \geqq \left(\dfrac{2}{n+2}\right)^{1/n}\left(\dfrac{n+2}{n}\right)^{\frac{1}{2}} & P \geqq 1-\left(\dfrac{2}{n+2}\right)^{2/n} \dfrac{1}{\lambda_0^2}, \\
(4) & \kappa \geqq 1, \ \left(\dfrac{2}{n+2}\right)^{1/n} \kappa \leqq \lambda_0 \leqq \kappa & & P \geqq 1-\dfrac{1}{\kappa^2}.
\end{array}$$

In addition, we know that for $\lambda_0 \geqq \kappa$, $\lambda_0 \geqq 1 : P \geqq 1-\dfrac{1}{\lambda_0^2}$. By rearranging these inequalities, we can bring them into the form in which they were given in the theorem, which is therefore proved.

It may be remarked that κ depends on the ratios between any two of the quantities $\lambda_1, \ldots, \lambda_n$, but it remains the same if $\lambda_1, \ldots, \lambda_n$ change in the same proportion.

Let us consider the sets of inequalities I, II and III of the theorem separately. The most important case in which set I is relevant occurs when nothing about the frequency distribution is known except the averages and standard deviations of the variables, in which case we have to put $\kappa = 0$. It provides a generalization of Tchebycheff's inequality which is obtained in the special case $n = 1$ in which $\lambda_0 = \lambda$.

If $n = 2$, the inequality refers to the frequency of all points lying inside the ellipse which has the axes $\lambda_1 \sqrt{2}\,\sigma_1$, $\lambda_2 \sqrt{2}\,\sigma_2$, and it can be written in the following way:

$$\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2} \leqq 2 : P \geqq 1-\frac{1}{2}\left(\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2}\right).$$

It is interesting to compare this limit with the one given by Berge (1937) for a rectangle which has its corners in the points with the co-ordinates $(\pm\lambda\sigma_1, \pm\lambda\sigma_2)$:

$$P \geqq 1-\frac{1+\sqrt{(1-r^2)}}{\lambda^2},$$

where r is the correlation coefficient of the frequency distribution. If $r = 0$, Berge's limit equals $1-2/\lambda^2$ and is therefore equal to the limit we get for the

ellipse with the axes $\lambda\sigma_1$, $\lambda\sigma_2$ which is inscribed to the rectangle. On the other hand, if $r = \pm 1$, Berge's limit becomes equal to $1-1/\lambda^2$, and we have to choose the circumscribed ellipse with the axes $\lambda\sqrt{2}\sigma_1$, $\lambda\sqrt{2}\sigma_2$ to obtain the same limit. Hence, the limit given here is certainly not inferior to the limit given by Berge, if the two variates are independent, but it is not superior to this limit, if there is a perfect correlation.

Set II is perhaps of more theoretical than practical interest. An intermediate value of κ may, however, be realized, if there is sufficient information about the frequencies inside, but not outside, a certain n-dimensional interval. The inequalities correspond to those given by Narumi (1923) for frequency distributions of one variate only, but having a more general meaning in so far as P may refer to multiples of other quantities besides the standard deviation. The first inequality seems the most interesting one of this set; it reduces to a special case of Narumi's inequality if we put $n = 1$:

$$\lambda \leqq \tfrac{2}{3}\kappa : P \geqq \frac{3}{2}\left(1-\frac{1}{\kappa^2}\right)\frac{\lambda}{\kappa},$$

and for $n = 2$ it may be written

$$\lambda_0 \leqq \frac{\kappa}{\sqrt{2}} : P \geqq 2\left(1-\frac{1}{\kappa^2}\right)\left(\frac{\lambda_0}{\kappa}\right)^2,$$

or

$$\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2} \geqq \frac{4}{\kappa^2} : P \geqq 4\left(1-\frac{1}{\kappa^2}\right)\frac{1}{\kappa^2(1/\lambda_1^2+1/\lambda_2^2)}.$$

Set III generalizes Gauss's inequalities. It reduces to the first two inequalities for those values of $\lambda_1, \ldots, \lambda_n$ (if any) for which $\kappa = \infty$. This is the case for all values of $\lambda_1, \ldots, \lambda_n$, if $y(hx_1, \ldots, hx_n)+y(-hx_1, \ldots, -hx_n)$ is a non-increasing function of $|h|$ for any fixed values of $x_1, \ldots, x_n$; especially if the function y decreases monotonically along every straight line radiating from the origin.

The assumption $\kappa = \infty$ will be made throughout the following analysis. Gauss's formulae are obtained by substituting $n = 1$:

$$\lambda \leqq \frac{2}{\sqrt{3}} : P \geqq \frac{\lambda}{\sqrt{3}},$$

$$\lambda \geqq \frac{2}{\sqrt{3}} : P \geqq 1-\frac{4}{9\lambda^2}.$$

In the case $n = 2$, the inequalities are also greatly simplified:

$$\lambda_0 \leqq 1 : P \geqq \frac{\lambda_0^2}{2},$$

$$\lambda_0 \geqq 1 : P \geqq 1-\frac{1}{2\lambda_0^2},$$

or

$$\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2} \geqq 2 : P \geqq \frac{1}{1/\lambda_1^2+1/\lambda_2^2},$$

$$\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2} \leqq 2 : P \geqq 1-\frac{1}{4}\left(\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2}\right).$$

Returning to the case of an unspecified value of n, we shall often be interested in a limit for the frequency of all values of the standardized or original variables for which the distance from the origin is no more than a certain value, i.e. either

$$\frac{x_1^2}{\sigma_1^2}+\ldots+\frac{x_n^2}{\sigma_n^2}\leqq\mu^2,$$

or

$$x_1^2+\ldots+x_n^2\leqq\rho^2.$$

In the first problem, we have

$$\lambda_1=\ldots=\lambda_n=\lambda_0=\frac{\mu}{\sqrt{n}},$$

and our inequalities can therefore be written

$$\mu^2\leqq 2^{2/n}(n+2)^{(n-2)/n}:P\geqq\frac{\mu^n}{(n+2)^{\frac{1}{2}n}},$$

$$\mu^2\geqq 2^{2/n}(n+2)^{(n-2)/n}:P\geqq 1-\left(\frac{2}{n+2}\right)^{2/n}\frac{n}{\mu^2}.$$

In the second problem

$$\lambda_1\sigma_1=\ldots=\lambda_n\sigma_n=\frac{\rho}{\sqrt{n}},\quad \lambda_0=\frac{\rho}{\sqrt{(\sigma_1^2+\ldots+\sigma_n^2)}}.$$

Hence, if

$$\rho^2\leqq 2^{2/n}(n+2)^{(n-2)/n}\frac{\sigma_1^2+\ldots+\sigma_n^2}{n}:P\geqq\left[\frac{n\rho^2}{(n+2)(\sigma_1^2+\ldots+\sigma_n^2)}\right]^{\frac{1}{2}n},$$

$$\rho^2\geqq 2^{2/n}(n+2)^{(n-2)/n}\frac{\sigma_1^2+\ldots+\sigma_n^2}{n}:P\geqq 1-\left(\frac{2}{n+2}\right)^{2/n}\frac{\sigma_1^2+\ldots+\sigma_n^2}{\rho^2}.$$

Again, the insertion of special values for n will simplify the formulae.

Finally, it may be at least of theoretical interest to consider the generalized Gauss limit, as obtained for $\kappa=\infty$, as a function of n for fixed values of λ_0, and to compare it with the generalized Tchebycheff limit which is obtained for $\kappa\leqq 1$ and is independent of n. For some particular values of λ_0 and n, this is done in the following table:

n \ λ_0	1	2	3
		Gauss limit	
1	0·577	0·889	0·951
2	0·500	0·875	0·944
3	0·457	0·864	0·940
4	0·423	0·856	0·936
		Tchebycheff limit	
Any value	0	0·750	0·889

Furthermore, since for $\lambda_0 \leqq \left(\frac{2}{n+2}\right)^{1/n} \left(\frac{n+2}{n}\right)^{\frac{1}{2}}$,

$$\lim_{n\to\infty} \left(\frac{n}{n+2}\right)^{\frac{1}{2}n} \lambda_0^n \leqq \lim_{n\to\infty} \frac{2}{n+2} = 0,$$

and since
$$\lim_{n\to\infty} \left[1-\left(\frac{2}{n+2}\right)^{2/n} \frac{1}{\lambda_0^2}\right] = 1-\frac{1}{\lambda_0^2},$$

it is seen that with an increasing number of variates, the Gauss limit loses gradually its superiority over the Tchebycheff limit and the difference between the two limits tends to vanish.

REFERENCES

BERGE, P. O. (1937). A note on a form of Tchebycheff's theorem for two variables. *Biometrika*, **29**, 405–6.

CAMP, B. H. (1922). A new generalization of Tchebycheff's statistical inequality. *Bull. Amer. Math. Soc.* **28**, 427–32.

GAUSS (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*. Göttingen, pp. 9–12.

NARUMI, S. (1923). On further inequalities with possible applications to problems in the theory of probability. *Biometrika*, **15**, 245–53.

PEARSON, K. (1919). On generalized Tchebycheff theorems in the mathematical theory of statistics. *Biometrika*, **12**, 284–96.

SMITH, C. D. (1930). On generalized Tchebycheff inequalities in mathematical statistics. *Amer. J. Math.* **52**, 109–26.

TCHEBYCHEFF (1867). Des valeurs moyennes. *J. Math. pures appl.* ser. 2, **12**, 177–84.

THE MODE AND MEDIAN OF A NEARLY NORMAL DISTRIBUTION WITH GIVEN CUMULANTS

By J. B. S. HALDANE, F.R.S.

In the early years of biometry the mode and median of a distribution, and especially the latter, were regarded as being almost as important as the mean. Later Pearson and others developed the method of moments, and recently the cumulants, which are readily derived from the moments, have been widely used. Pearson (1895 and after) discussed the relation of the mean, mode and median in some of the skew frequency curves which he invented. He found empirically that for skew curves of Type III, namely,

$$y = y_0\left(1+\frac{x}{a}\right)^p e^{-px/a},$$

when p is positive, Mode − Mean = 3 (Median − Mean) approximately. By fitting for a series of integral values of p he found

$$\frac{\text{Median}-\text{Mode}}{\text{Mean}-\text{Mode}} = 0{\cdot}6691 + 0{\cdot}0094p^{-1}.$$

Some later writers have taken the trisection as a general law. But as we shall see, it does not always hold, even approximately. So far as I know, general expressions for the mode and median in terms of the cumulants have not been given, nor have the conditions been stated under which Pearson's rule holds approximately.

Consider a variate X, with distribution $df = F(X)\,dX$, and cumulants $\kappa_1 = m$, $\kappa_2 = \sigma^2$, $\kappa_3, \ldots, \kappa_r, \ldots$.

The algebra is simplified if we make the transformation $x = (X-m)/\sigma$, so that x has mean zero, and unit standard deviation, its cumulants being $0, 1, \gamma_1, \gamma_2, \ldots, \gamma_r, \ldots$, where $\gamma_{r-2} = \kappa_r \kappa_2^{-\frac{1}{2}r}$. Thus γ_r is the rth measure of the deviation from normality of the distribution of X. Now γ_r may become infinite for all values of r, or for all even values of r, above a certain value. It may diminish indefinitely or remain small. In other cases it falls with r at first, but then increases without limit. Thus for Pearson's Type III distribution, whose equation has been given above,

$$\kappa_r = (r-1)!\,(p+1)\left(\frac{a}{p}\right)^r, \quad \text{and} \quad \gamma_r = (r+1)!\,(p+1)^{-\frac{1}{2}r}.$$

And in the case of any estimate of a statistic, such as the mean, variance, standard deviation, or skewness, based on a sample of n members, $\gamma_r = O(n^{-\frac{1}{2}r})$. We shall not discuss the convergence of the series obtained later. But it is worth

noting that expansions in Hermitian polynomials are often satisfactory asymptotic expansions, even if they diverge.

Let $df = f(x)\,dx$ be the distribution of x. Then it is known that we may write, symbolically,*

$$f(x) = \exp\left[-\frac{\gamma_1}{3!}\left(\frac{d}{dx}\right)^3 + \frac{\gamma_2}{4!}\left(\frac{d}{dx}\right)^4 - \dots + \frac{\gamma_{r-2}}{r!}\left(\frac{-d}{dx}\right)^r \dots\right]\frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}.$$

Expanding in Hermitian polynomials $H_r(x) = e^{\frac{1}{2}x^2}\frac{d^r}{dx^r}e^{-\frac{1}{2}x^2}$, we have

$$f(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}\left[1 - \frac{\gamma_1}{3!}H_3(x) + \frac{\gamma_2}{4!}H_4(x) - \frac{\gamma_3}{5!}H_5(x) + \frac{10\gamma_1^2+\gamma_4}{6!}H_6(x) - \frac{35\gamma_1\gamma_2+\gamma_5}{7!}H_7(x) + \dots\right].$$

In the special case where $\gamma_r = O(n^{-\frac{1}{2}r})$ we have

$$f(x) = \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}\left[1 - \frac{\gamma_1}{6}H_3(x) + \frac{\gamma_2}{24}H_4(x) - \frac{\gamma_3}{120}H_5(x) + \frac{\gamma_1^2}{72}H_6(x) - \frac{\gamma_1\gamma_2}{144}H_7(x) - \frac{\gamma_1^3}{1296}H_9(x) + O(n^{-2})\right].$$

To find the mode, we put $\frac{d}{dx}f(x) = 0$. Hence

$$H_1(x) - \frac{\gamma_1}{6}H_4(x) + \frac{\gamma_2}{24}H_5(x) + \dots = 0.$$

The terms needed to find the root of this equation with the smallest absolute value depend, of course, on the behaviour of the cumulants as r increases. If $|\gamma_2|$, $|\gamma_3|$, etc., are not substantially less than $|\gamma_1|$, but all are small, so that powers and products can be neglected, then:

$$x = -\frac{\gamma_1}{2} + \frac{\gamma_3}{8} - \frac{\gamma_5}{48} + \frac{\gamma_7}{576} - \dots.$$

Where $\gamma_r = O(n^{-\frac{1}{2}r})$, we have

$$\gamma_1 x^2 - \left(1 + \frac{5\gamma_2}{8} - \frac{35\gamma_1^2}{24}\right)x - \frac{\gamma_1}{2} + \frac{\gamma_3}{8} - \frac{35\gamma_1\gamma_2}{48} + \frac{35\gamma_1^3}{48} + O(n^{-\frac{5}{2}}) = 0.$$

Hence $$x = -\frac{\gamma_1}{2} + \frac{\gamma_3}{8} - \frac{5\gamma_1\gamma_2}{12} + \frac{\gamma_1^3}{4} + O(n^{-2}).$$

So in the first case the mode is

$$m + \sigma\left(-\frac{\gamma_1}{2} + \frac{\gamma_3}{8} - \frac{\gamma_5}{48} + \dots + \frac{(-)^r\gamma_{2r-1}}{2^r r!} + \dots\right)$$
$$= \kappa_1 - \frac{\kappa_3}{2\kappa_2} + \frac{\kappa_5}{8\kappa_2^2} - \frac{\kappa_7}{48\kappa_2^3} + \dots + \frac{(-)^r\kappa_{2r+1}}{2^r r!\,\kappa_2^r} + \dots. \qquad (1)$$

* For a comparison of this expansion with that of the Charlier Type A distribution, see Cramér (1937). The symbolic relation has also been used recently by Cornish & Fisher (1937).

If $\kappa_r = O(n)$, then the mode is

$$m - \tfrac{1}{2}\sigma\left(\gamma_1 - \frac{\gamma_3}{4} + \frac{5\gamma_1\gamma_2}{6} - \frac{\gamma_1^3}{2}\right) + O(n^{-2})$$

$$= \kappa_1 - \frac{\kappa_3}{2\kappa_2} + \frac{\kappa_5}{8\kappa_2^2} - \frac{5\kappa_3\kappa_4}{12\kappa_2^3} + \frac{\kappa_3^3}{4\kappa_2^4} + O(n^{-2})$$

$$= \mu_1' - \frac{\mu_3}{2\mu_2} + \frac{\mu_5}{8\mu_2^2} - \frac{5\mu_3\mu_4}{12\mu_2^3} + \frac{\mu_3^3}{4\mu_2^4} + O(n^{-2}) \ldots. \tag{2}$$

The median is the value of x for which $\int_{-\infty}^{x} f(u)\,du = \frac{1}{2}$, or, since

$$\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{0} e^{-\frac{1}{2}u^2}\,du = \tfrac{1}{2},$$

the value for which

$$\frac{1}{\sqrt{(2\pi)}}\int_{0}^{x} e^{-\frac{1}{2}u^2}\,du + \frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}\left[-\frac{\gamma_1}{6}H_2(x) + \frac{\gamma_2}{24}H_3(x) + \ldots\right] = 0.$$

So $$e^{\frac{1}{2}x^2}\int_{0}^{x} e^{-\frac{1}{2}u^2}\,du - \frac{\gamma_1}{6}H_2(x) + \frac{\gamma_2}{24}H_3(x) - \frac{\gamma_3}{120}H_4(x) + \ldots = 0.$$

If γ_r does not fall off systematically we have

$$x = -\frac{\gamma_1}{6} + \frac{\gamma_3}{40} - \frac{\gamma_5}{336} + \ldots + \frac{(-)^r\gamma_{2r-1}}{2^r(2r+1)\,r!} + \ldots.$$

When $\kappa_r = O(n)$, we find

$$\frac{x^3}{3} - \frac{\gamma_1 x^2}{6} + \left(1 + \frac{\gamma_2}{8} - \frac{5\gamma_1^2}{4}\right)x + \frac{\gamma_1}{6} - \frac{\gamma_3}{40} + \frac{5\gamma_1\gamma_2}{48} - \frac{35\gamma_1^3}{432} + O(n^{-\frac{3}{2}}) = 0.$$

Hence $$x = -\frac{\gamma_1}{6} + \frac{\gamma_3}{40} - \frac{\gamma_1\gamma_2}{12} + \frac{17\gamma_1^3}{324} + O(n^{-\frac{3}{2}}).$$

So if powers and products can be neglected the median is

$$m + \sigma\left(-\frac{\gamma_1}{6} + \frac{\gamma_3}{40} - \frac{\gamma_5}{336} + \ldots\right)$$

$$= \kappa_1 - \frac{\kappa_3}{6\kappa_2} + \frac{\kappa_5}{40\kappa_2^2} - \frac{\kappa_7}{336\kappa_2^3} + \ldots + \frac{(-)^r\kappa_{2r+1}}{2^r(2r+1)\,r!\,\kappa_2^r} + \ldots. \tag{3}$$

And when $\kappa_r = O(n)$, the median is

$$m - \tfrac{1}{2}\sigma\left(\frac{\gamma_1}{3} - \frac{\gamma_3}{20} + \frac{\gamma_1\gamma_2}{6} - \frac{17\gamma_1^3}{162}\right) + O(n^{-2})$$

$$= \kappa_1 - \frac{\kappa_3}{6\kappa_2} + \frac{\kappa_5}{40\kappa_2^2} - \frac{\kappa_3\kappa_4}{12\kappa_2^3} + \frac{17\kappa_3^3}{324\kappa_2^4} - O(n^{-2}) \ldots$$

$$= \mu_1' - \frac{\mu_3}{6\mu_2} + \frac{\mu_5}{40\mu_2^2} - \frac{\mu_3\mu_4}{12\mu_2^3} + \frac{17\mu_3^3}{324\mu_2^4} - O(n^{-2}). \tag{4}$$

Thus the distance from the mean to the mode is approximately thrice that from the mean to the median if γ_2, γ_4, etc., are small quantities (i.e. β_2 is nearly 3, etc.) and if γ_3, γ_5, etc., are small compared with γ_1. This is often the case for nearly normal distributions. Thus for the best known of Type III distributions, that of χ^2 for n degrees of freedom, $\kappa_1 = n$, $\kappa_2 = 2n$, $\kappa_r = 2^{r-1}(r-1)!\,n$. Hence from equation (2) the mode is at $n-2+O(n^{-2})$. In fact it is $n-2$ exactly. For the median, equation (4) gives $n-\frac{2}{3}+\frac{32}{405n}-O(n^{-2})$. The fraction which Pearson empirically estimated at $0{\cdot}6691+0{\cdot}0094p^{-1}$ is therefore $\frac{2}{3}+\frac{8}{405p}+O(p^{-2})$.

Consider the distribution of the estimated mean, $n^{-1}\Sigma x_r$, from a sample of n, taken from a distribution with arbitrary cumulants κ_1, κ_2, κ_3, etc. The cumulants of the distribution of the mean are κ_1, κ_2/n, κ_3/n^2, etc. Hence equations (2) and (4) hold, and the mean is κ_1, the mode is

$$\kappa_1-\frac{\kappa_3}{2n\kappa_2}+\left(\frac{\kappa_5}{8\kappa_2^2}-\frac{5\kappa_3\kappa_4}{12\kappa_2^3}+\frac{\kappa_3^3}{4\kappa_2^4}\right)n^{-2}+O(n^{-3}),$$

and the median is

$$\kappa_1-\frac{\kappa_3}{6n\kappa_2}+\left(\frac{\kappa_5}{40\kappa_2^2}-\frac{\kappa_3\kappa_4}{12\kappa_2^3}+\frac{17\kappa_3^3}{324\kappa_2^4}\right)n^{-2}+O(n^{-3}).$$

The corresponding expressions in terms of the moments may be written down. The skewness is not exactly $1/n$ of the skewness of the original curve, but nearly so if n is large.

Again consider the distribution of the estimated variance from a sample of n members from a normal distribution with unknown mean, and variance σ^2. The cumulants of the distribution of the estimated variance $\frac{1}{n-1}[\Sigma x_r^2-n^{-1}(\Sigma x_r)^2]$ are

$$\kappa_1=\sigma_1^2,\quad \kappa_2=\frac{2}{n-1},\quad \kappa_3=\frac{8\sigma^6}{(n-1)^2},\quad \kappa_4=\frac{48\sigma^8}{(n-1)^3},\quad \kappa_5=\frac{384\sigma^{10}}{(n-1)^4},\quad \text{etc.}$$

Thus we are dealing with a slightly transformed χ^2 distribution, and the mode is $\sigma^2-\frac{2\sigma^2}{n-1}$, or $\frac{(n-3)\sigma^2}{n-1}$, exactly. From equation (4) the median is

$$\sigma^2\left[1-\frac{2}{3(n-1)}+\frac{32}{405(n-1)^2}+O(n^{-3})\right],\quad \text{or}\quad \sigma^2\left[1-\frac{2}{3n}-\frac{302}{405n^2}+O(n^{-3})\right].$$

If, however, the distribution sampled is not normal, but has a finite κ_3, etc., then the first three cumulants, in terms of those of the original distribution, are

$$\kappa_2,\quad \frac{2\kappa_2^2}{n-1}+\frac{\kappa_4}{n}\quad \text{and}\quad \frac{8\kappa_2^3}{(n-1)^2}+\frac{4(n-2)\kappa_3^2}{n(n-1)^2}+\frac{12\kappa_2\kappa_4}{n(n-1)}+\frac{\kappa_6}{n^2}.$$

Fisher (1928) gives expressions for the next three cumulants, but these become very complex, $\kappa(2^6)$ having 21 terms.

It follows that the mean is κ_2, the mode

$$\kappa_2 - \left[2\kappa_2 + \frac{4\kappa_3^2 + 8\kappa_2\kappa_4 + \kappa_6}{2(2\kappa_2^2 + \kappa_4)}\right] n^{-1} + O(n^{-2}),$$

and the median

$$\kappa_2 - \left[\frac{2\kappa_2}{3} + \frac{4\kappa_3^2 + 8\kappa_2\kappa_4 + \kappa_6}{6(2\kappa_2^2 + \kappa_4)}\right] n^{-1} + O(n^{-2}).$$

The distribution of the estimate of κ_3 is symmetrical for a symmetrical distribution. In general its mean is κ_3, its mode

$$\kappa_3 - \left[45\kappa_3 + \frac{-558\kappa_3^3 + 216\kappa_2^2\kappa_5 + 108\kappa_4\kappa_5 - 9\kappa_3\kappa_6 + 27\kappa_2\kappa_7 + \kappa_9}{2(6\kappa_2^3 + 9\kappa_3^2 + 9\kappa_2\kappa_4 + \kappa_6)}\right] n^{-1} + O(n^{-2}),$$

the median differing from the mean by $\frac{1}{3}$ of the value given above.

Finally the mean estimate of κ_4 for a sample from a normal distribution is zero, its mode $-\dfrac{36\sigma^4}{n} + O(n^{-2})$, and its median $-\dfrac{12\sigma^4}{n} + O(n^{-2})$. The general expressions for the mode and median can readily be calculated from Fisher's (1928) equations.

We now pass to some cases where Pearson's trisection rule does not hold, even approximately.

Pearson's Type I and Type IV curves are asymmetrical, and have one more adjustable parameter than Type III. Thus γ_2, i.e. $\beta_2 - 3$, can vary independently of γ_1, i.e. $\beta_1^{\frac{1}{2}}$. In consequence the curves may be nearly symmetrical, but far from normal. This is so if they approximate to Type II or Type VII, respectively. In this case the even measures of divergence γ_{2r} may be much larger than the odd measures γ_{2r+1}, which tend to zero with symmetry. Thus we cannot neglect higher cumulants, or products, as in equations 1–4. For Type IV and VII curves the higher moments are infinite, so no formulae of the given type are possible. For Types I and II a formula could be given, but direct calculation is clearly preferable.

A simpler case is that of the scalene triangular distribution whose graph is obtained by joining $(-b, 0)$, $\left(0, \dfrac{2}{a+b}\right)$, and $(a, 0)$. That is to say

$$y = \frac{2(b+x)}{b(a+b)} \quad \text{for} \quad b \leqslant x \leqslant 0, \qquad \text{and} \qquad y = \frac{2(a-x)}{a(a+b)} \quad \text{for} \quad 0 \leqslant x \leqslant a.$$

Here the mean is $\frac{1}{3}(a-b)$, the mode O, and the median, if $a > b$, is

$$a - \sqrt{[\tfrac{1}{2}a(a+b)]}, \quad \text{or} \quad \frac{a-b}{4} + \frac{(a-b)^2}{32a} + \frac{(a-b)^3}{128a^2} + \dots.$$

That is to say, the median is one-quarter, not one-third, of the distance from the mean to the mode, if $a-b$ is small compared with a. The rth moment about zero is $\dfrac{2[a^{r+1} - (-b)^{r+1}]}{r(r+1)(a+b)}$

If $k = \frac{a-b}{\sqrt{(ab)}}$, then $\gamma_1 = \frac{\sqrt{[2k(9+2k_3^2)]}}{5(3+k^2)^{\frac{3}{2}}}\gamma_2 = -\frac{3}{5}$, $\gamma_3 = -\frac{6\sqrt{[2k(9+2k^2)]}}{7(3+k^2)^{\frac{3}{2}}}$, etc.
If k is small, the odd γ's are of order k, whilst the even ones approximate to those of the symmetrical distribution. Moreover $\gamma_3 = -\frac{30}{7}\gamma_1$, so the formulae 1–4 clearly do not hold.

We see then that formulae (2) and (4) are quite useful in a special case which is important in sampling theory, but have no general validity.

SUMMARY

Expressions are obtained by which the distances of the mean and median of a skew distribution can be calculated from its moments or cumulants. The series obtained may or may not converge. They give satisfactory results for Type III distributions, and for the distributions of the mean, variance, and other cumulants as estimated from samples.

REFERENCES

CORNISH, E. A. & FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.* Pt. 4.

CRAMÉR, H. (1937). *Random Variables and Probability Distributions*, Cambridge Mathematical Tracts, 86–88.

FISHER, R. A. (1928). Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc.* Ser. 2, **30**, 199.

PEARSON, K. (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philos. Trans.* **186**, 343.

TABLE OF PERCENTAGE POINTS OF THE t-DISTRIBUTION

COMPUTED BY MAXINE MERRINGTON

THE following table has been derived from Miss Thompson's Tables of Percentage Points of the Incomplete Beta Function (*Biometrika*, **32**, 168–81), by taking

$$t = \sqrt{\{\nu_2(1-x)/\nu_1 x\}}$$

for the case $\nu_1 = 1$. t is the usual "Student" ratio based on an estimate of variance having $\nu = \nu_2$ degrees of freedom. If t_P is the quantity tabled, $P/100$ is the chance that $|t| \geqslant t_P$, i.e. represents the area in the two tails of the t-distribution. The table includes certain levels for t not previously available and should be accurate to the five significant figures shown.

ν \ P	50	25	10	5	2·5	1	0·5
1	1·00000	2·4142	6·3138	12·706	25·452	63·657	127·32
2	0·81650	1·6036	2·9200	4·3027	6·2053	9·9248	14·089
3	0·76489	1·4226	2·3534	3·1825	4·1765	5·8409	7·4533
4	0·74070	1·3444	2·1318	2·7764	3·4954	4·6041	5·5976
5	0·72669	1·3009	2·0150	2·5706	3·1634	4·0321	4·7733
6	0·71756	1·2733	1·9432	2·4469	2·9687	3·7074	4·3168
7	0·71114	1·2543	1·8946	2·3646	2·8412	3·4995	4·0293
8	0·70639	1·2403	1·8595	2·3060	2·7515	3·3554	3·8325
9	0·70272	1·2297	1·8331	2·2622	2·6850	3·2498	3·6897
10	0·69981	1·2213	1·8125	2·2281	2·6338	3·1693	3·5814
11	0·69745	1·2145	1·7959	2·2010	2·5931	3·1058	3·4966
12	0·69548	1·2089	1·7823	2·1788	2·5600	3·0545	3·4284
13	0·69384	1·2041	1·7709	2·1604	2·5326	3·0123	3·3725
14	0·69242	1·2001	1·7613	2·1448	2·5096	2·9768	3·3257
15	0·69120	1·1967	1·7530	2·1315	2·4899	2·9467	3·2860
16	0·69013	1·1937	1·7459	2·1199	2·4729	2·9208	3·2520
17	0·68919	1·1910	1·7396	2·1098	2·4581	2·8982	3·2225
18	0·68837	1·1887	1·7341	2·1009	2·4450	2·8784	3·1966
19	0·68763	1·1866	1·7291	2·0930	2·4334	2·8609	3·1737
20	0·68696	1·1848	1·7247	2·0860	2·4231	2·8453	3·1534
21	0·68635	1·1831	1·7207	2·0796	2·4138	2·8314	3·1352
22	0·68580	1·1816	1·7171	2·0739	2·4055	2·8188	3·1188
23	0·68531	1·1802	1·7139	2·0687	2·3979	2·8073	3·1040
24	0·68485	1·1789	1·7109	2·0639	2·3910	2·7969	3·0905
25	0·68443	1·1777	1·7081	2·0595	2·3846	2·7874	3·0782
26	0·68405	1·1766	1·7056	2·0555	2·3788	2·7787	3·0669
27	0·68370	1·1757	1·7033	2·0518	2·3734	2·7707	3·0565
28	0·68335	1·1748	1·7011	2·0484	2·3685	2·7633	3·0469
29	0·68304	1·1739	1·6991	2·0452	2·3638	2·7564	3·0380
30	0·68276	1·1731	1·6973	2·0423	2·3596	2·7500	3·0298
40	0·68066	1·1673	1·6839	2·0211	2·3289	2·7045	2·9712
60	0·67862	1·1616	1·6707	2·0003	2·2991	2·6603	2·9146
120	0·67656	1·1559	1·6577	1·9799	2·2699	2·6174	2·8599
∞	0·67449	1·1503	1·6449	1·9600	2·2414	2·5758	2·8070

THE PROBABILITY INTEGRAL OF THE RANGE IN SAMPLES OF n OBSERVATIONS FROM A NORMAL POPULATION

I. FOREWORD AND TABLES

By E. S. PEARSON

1. *Scope of the main table*

Denote by $x_1, x_2, \ldots, x_n$ a random sample of n observations, arranged in ascending order of magnitude, drawn from a normal or Gaussian population having for probability law

$$p(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left[-\tfrac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \tag{1}$$

where μ and σ are respectively the mean and standard deviation of the population. The range, sometimes described as the spread, in the sample is $x_n - x_1$ and we shall write the ratio of the range to the population standard deviation as

$$w = \frac{x_n - x_1}{\sigma} \tag{2}$$

No simple expression exists for the probability law $f_n(w)$ of w, but Table 1 below gives for specific values W of w computed values of the probability *integral*

$$P_n(W) = \int_0^W f_n(w)\,dw \tag{3}$$

This expression is the chance that the range in a sample of n observations is less than a given multiple of the population standard deviation. The table has been calculated for samples with n lying between 2 and 20 and for intervals of 0·05 of W. The values of the integral are given to 4-decimal place accuracy; linear interpolation is adequate except in the neighbourhood of the two quartiles in each column. The method of calculation is described below by Dr H. O. Hartley in a separate section.

2. *Auxiliary table for special uses* (Table 2)

When dealing with samples containing only a small number of observations the range or spread may often be usefully employed as a measure of dispersion in place either of the standard (root mean square) deviation or the mean deviation. For example, an estimate of the population standard deviation may be obtained by multiplying the range in a single sample or the mean range in a number of samples by the factors a_n shown in Table 2. The accuracy of this form of estimation of σ compared with that of other methods has been discussed by Davies & Pearson (1934).

In other circumstances it may be useful to plot in serial order on a control chart the values of range obtained from successive samples, e.g. when dealing with the control of quality in mass production. For this purpose it is necessary to know certain standard probability levels for w which will serve as control limits*. Twelve of these levels expressed as percentage points and obtained by interpolation in the main Table 1, are shown in Table 2. They replace approximate limits published a few years ago (Pearson, 1932). It will be found, however, that except in the case $n = 3$†, the discrepancies between the two tables are all small. As an example, the table shows that if samples of 7 observations are randomly drawn from a population with a standard deviation σ, then in the long run only 5 % of these should have a range greater than $4{\cdot}17\sigma$, while 95 % should satisfy the inequality

$$1{\cdot}25\sigma \leqslant x_n - x_1 \leqslant 4{\cdot}49\sigma$$

Probability levels for samples with $n > 12$ have not been included as the use of range for control purposes in larger samples is of doubtful value.

* For a discussion of the use of range in problems of industrial quality control see Reports issued by the British Standards Institution (Pearson, 1935, pp. 89–90; Dudding & Jennett, 1942).

† Correct values for the case $n=3$ were given by McKay & Pearson (1933).

Table 1. *Probability integral of the range W in normal samples of size n*

W \ n	2	3	4	5	6	7	8	9	10
0·00	0·0000	0·0000							
0·05	·0282	·0007	0·0000						
0·10	·0564	·0028	·0001						
0·15	·0845	·0062	·0004	0·0000					
0·20	·1125	·0110	·0010	·0001					
0·25	0·1403	0·0171	0·0020	0·0002					
0·30	·1680	·0245	·0034	·0004	0·0000				
0·35	·1955	·0332	·0053	·0008	·0001				
0·40	·2227	·0431	·0079	·0014	·0002	0·0000			
0·45	·2497	·0543	·0111	·0022	·0004	·0001			
0·50	0·2763	0·0666	0·0152	0·0033	0·0007	0·0002	0·0000		
0·55	·3027	·0800	·0200	·0048	·0011	·0003	·0001		
0·60	·3286	·0944	·0257	·0068	·0017	·0004	·0001	0·0000	
0·65	·3542	·1099	·0323	·0092	·0025	·0007	·0002	·0001	
0·70	·3794	·1263	·0398	·0121	·0036	·0011	·0003	·0001	0·0000
0·75	0·4041	0·1436	0·0483	0·0157	0·0050	0·0016	0·0005	0·0002	0·0001
0·80	·4284	·1616	·0578	·0200	·0068	·0023	·0008	·0002	·0001
0·85	·4522	·1805	·0682	·0250	·0090	·0032	·0011	·0004	·0001
0·90	·4755	·2000	·0797	·0309	·0117	·0044	·0016	·0006	·0002
0·95	·4983	·2201	·0922	·0375	·0150	·0059	·0023	·0009	·0003
1·00	0·5205	0·2407	0·1057	0·0450	0·0188	0·0078	0·0032	0·0013	0·0005
1·05	·5422	·2618	·1201	·0535	·0234	·0101	·0043	·0018	·0008
1·10	·5633	·2833	·1355	·0629	·0287	·0129	·0057	·0025	·0011
1·15	·5839	·3051	·1517	·0733	·0348	·0163	·0075	·0035	·0016
1·20	·6039	·3272	·1688	·0847	·0417	·0203	·0098	·0047	·0022
1·25	0·6232	0·3495	0·1868	0·0970	0·0495	0·0250	0·0125	0·0062	0·0030
1·30	·6420	·3719	·2054	·1104	·0583	·0304	·0157	·0080	·0041
1·35	·6602	·3943	·2248	·1247	·0680	·0366	·0195	·0103	·0054
1·40	·6778	·4168	·2448	·1400	·0787	·0437	·0240	·0131	·0071
1·45	·6948	·4392	·2654	·1562	·0904	·0517	·0292	·0164	·0092
1·50	0·7112	0·4614	0·2865	0·1733	0·1031	0·0606	0·0353	0·0204	0·0117
1·55	·7269	·4835	·3080	·1913	·1168	·0705	·0422	·0250	·0148
1·60	·7421	·5053	·3299	·2101	·1316	·0814	·0499	·0304	·0184
1·65	·7567	·5269	·3521	·2296	·1473	·0934	·0587	·0366	·0227
1·70	·7707	·5481	·3745	·2498	·1639	·1064	·0684	·0437	·0277
1·75	0·7841	0·5690	0·3971	0·2706	0·1815	0·1204	0·0792	0·0517	0·0336
1·80	·7969	·5894	·4197	·2920	·2000	·1355	·0910	·0607	·0403
1·85	·8092	·6094	·4423	·3138	·2193	·1516	·1039	·0707	·0479
1·90	·8209	·6290	·4649	·3361	·2394	·1686	·1178	·0818	·0565
1·95	·8321	·6480	·4874	·3587	·2602	·1867	·1329	·0940	·0661
2·00	0·8427	0·6665	0·5096	0·3816	0·2816	0·2056	0·1489	0·1072	0·0768
2·05	·8528	·6845	·5317	·4046	·3035	·2254	·1661	·1216	·0886
2·10	·8624	·7019	·5534	·4277	·3260	·2460	·1842	·1371	·1015
2·15	·8716	·7187	·5748	·4508	·3489	·2673	·2033	·1536	·1156
2·20	·8802	·7349	·5957	·4739	·3720	·2893	·2232	·1712	·1307
2·25	0·8884	0·7505	0·6163	0·4969	0·3955	0·3118	0·2440	0·1899	0·1470
2·30	·8961	·7655	·6363	·5196	·4190	·3348	·2656	·2095	·1645
2·35	·9034	·7799	·6558	·5421	·4427	·3582	·2878	·2300	·1830
2·40	·9103	·7937	·6748	·5643	·4663	·3820	·3107	·2514	·2025
2·45	·9168	·8069	·6932	·5861	·4899	·4059	·3341	·2735	·2230
2·50	0·9229	0·8195	0·7110	0·6075	0·5132	0·4300	0·3579	0·2964	0·2443

Table 1 (cont.). *Probability integral of the range W in normal samples of size n*

W \ n	11	12	13	14	15	16	17	18	19	20
0·85	0·0000									
0·90	·0001									
0·95	·0001	0·0000								
1·00	0·0002	0·0001	0·0000							
1·05	·0003	·0001	·0001							
1·10	·0005	·0002	·0001	0·0000						
1·15	·0007	·0003	·0001	·0001						
1·20	·0010	·0005	·0002	·0001	0·0000					
1·25	0·0015	0·0007	0·0004	0·0002	0·0001	0·0000				
1·30	·0021	·0010	·0005	·0003	·0001	·0001	0·0000			
1·35	·0028	·0015	·0008	·0004	·0002	·0001	·0001	0·0000		
1·40	·0038	·0021	·0011	·0006	·0003	·0002	·0001	·0001		
1·45	·0051	·0028	·0016	·0009	·0005	·0003	·0001	·0001	0·0000	
1·50	0·0067	0·0038	0·0022	0·0012	0·0007	0·0004	0·0002	0·0001	0·0001	0·0000
1·55	·0087	·0051	·0030	·0017	·0010	·0006	·0003	·0002	·0001	·0001
1·60	·0111	·0067	·0040	·0024	·0014	·0008	·0005	·0003	·0002	·0001
1·65	·0140	·0086	·0053	·0032	·0020	·0012	·0007	·0004	·0003	·0002
1·70	·0175	·0111	·0069	·0043	·0027	·0017	·0011	·0007	·0004	·0003
1·75	0·0217	0·0140	0·0090	0·0058	0·0037	0·0024	0·0015	0·0010	0·0006	0·0004
1·80	·0266	·0175	·0115	·0075	·0049	·0032	·0021	·0013	·0009	·0006
1·85	·0323	·0217	·0145	·0097	·0065	·0043	·0028	·0019	·0012	·0008
1·90	·0388	·0266	·0182	·0124	·0084	·0057	·0039	·0026	·0018	·0012
1·95	·0463	·0323	·0225	·0156	·0108	·0075	·0052	·0035	·0024	·0017
2·00	0·0548	0·0389	0·0276	0·0195	0·0137	0·0097	0·0068	0·0048	0·0033	0·0023
2·05	·0643	·0465	·0335	·0241	·0173	·0124	·0088	·0063	·0045	·0032
2·10	·0749	·0550	·0403	·0295	·0215	·0156	·0114	·0082	·0060	·0043
2·15	·0866	·0646	·0481	·0357	·0264	·0196	·0144	·0106	·0078	·0057
2·20	·0994	·0753	·0569	·0429	·0323	·0242	·0181	·0136	·0102	·0076
2·25	0·1134	0·0872	0·0669	0·0511	0·0390	0·0297	0·0226	0·0172	0·0130	0·0099
2·30	·1286	·1003	·0779	·0605	·0468	·0361	·0279	·0215	·0165	·0127
2·35	·1450	·1145	·0902	·0709	·0556	·0435	·0340	·0265	·0207	·0161
2·40	·1625	·1300	·1037	·0825	·0655	·0519	·0411	·0325	·0256	·0202
2·45	·1811	·1466	·1183	·0953	·0766	·0615	·0493	·0394	·0315	·0251
2·50	0·2007	0·1644	0·1342	0·1094	0·0890	0·0722	0·0586	0·0474	0·0383	0·0309

Table 1 (cont.). *Probability integral of the range W in normal samples of size n*

W \ n	2	3	4	5	6	7	8	9	10
2·50	0·9229	0·8195	0·7110	0·6075	0·5132	0·4300	0·3579	0·2964	0·2443
2·55	·9286	·8315	·7282	·6283	·5364	·4541	·3820	·3198	·2665
2·60	·9340	·8429	·7448	·6487	·5592	·4782	·4064	·3437	·2894
2·65	·9390	·8537	·7607	·6685	·5816	·5022	·4309	·3680	·3130
2·70	·9438	·8640	·7759	·6877	·6036	·5259	·4555	·3927	·3372
2·75	0·9482	0·8737	0·7905	0·7063	0·6252	0·5494	0·4801	0·4175	0·3617
2·80	·9523	·8828	·8045	·7242	·6461	·5725	·5044	·4425	·3867
2·85	·9561	·8915	·8177	·7415	·6665	·5952	·5286	·4675	·4119
2·90	·9597	·8996	·8304	·7580	·6863	·6174	·5525	·4923	·4372
2·95	·9630	·9073	·8424	·7739	·7055	·6390	·5760	·5171	·4625
3·00	0·9661	0·9145	0·8537	0·7891	0·7239	0·6601	0·5991	0·5415	0·4878
3·05	·9690	·9212	·8645	·8036	·7416	·6806	·6216	·5656	·5129
3·10	·9716	·9275	·8746	·8174	·7587	·7003	·6436	·5892	·5378
3·15	·9741	·9334	·8842	·8305	·7750	·7194	·6649	·6124	·5623
3·20	·9763	·9388	·8931	·8429	·7905	·7377	·6856	·6350	·5864
3·25	0·9784	0·9439	0·9016	0·8546	0·8053	0·7553	0·7055	0·6569	0·6099
3·30	·9804	·9487	·9095	·8657	·8194	·7721	·7248	·6782	·6329
3·35	·9822	·9531	·9168	·8761	·8327	·7881	·7432	·6988	·6553
3·40	·9838	·9572	·9237	·8859	·8454	·8034	·7609	·7186	·6769
3·45	·9853	·9609	·9302	·8951	·8573	·8179	·7778	·7376	·6978
3·50	0·9867	0·9644	0·9361	0·9037	0·8685	0·8316	0·7939	0·7558	0·7180
3·55	·9879	·9677	·9417	·9117	·8790	·8446	·8091	·7732	·7373
3·60	·9891	·9706	·9468	·9192	·8889	·8568	·8236	·7898	·7558
3·65	·9901	·9734	·9516	·9261	·8981	·8683	·8372	·8055	·7735
3·70	·9911	·9759	·9559	·9326	·9067	·8790	·8501	·8204	·7902
3·75	0·9920	0·9782	0·9600	0·9386	0·9148	0·8891	0·8622	0·8345	0·8062
3·80	·9928	·9803	·9637	·9441	·9222	·8985	·8736	·8477	·8212
3·85	·9935	·9822	·9672	·9493	·9291	·9073	·8842	·8602	·8355
3·90	·9942	·9839	·9703	·9540	·9355	·9155	·8941	·8718	·8488
3·95	·9948	·9856	·9732	·9583	·9415	·9230	·9034	·8827	·8614
4·00	0·9953	0·9870	0·9758	0·9623	0·9469	0·9300	0·9120	0·8929	0·8731
4·05	·9958	·9883	·9782	·9660	·9519	·9365	·9199	·9024	·8841
4·10	·9963	·9895	·9804	·9693	·9566	·9425	·9273	·9112	·8943
4·15	·9967	·9906	·9824	·9724	·9608	·9480	·9341	·9193	·9038
4·20	·9970	·9916	·9842	·9752	·9647	·9530	·9404	·9269	·9126
4·25	0·9974	0·9925	0·9859	0·9777	0·9682	0·9576	0·9461	0·9338	0·9208
4·30	·9976	·9933	·9874	·9800	·9715	·9619	·9514	·9402	·9283
4·35	·9979	·9941	·9887	·9821	·9744	·9657	·9562	·9460	·9352
4·40	·9981	·9947	·9899	·9840	·9771	·9692	·9607	·9514	·9416
4·45	·9984	·9953	·9910	·9857	·9795	·9724	·9647	·9563	·9474
4·50	0·9985	0·9958	0·9920	0·9873	0·9817	0·9754	0·9684	0·9608	0·9527
4·55	·9987	·9963	·9929	·9887	·9837	·9780	·9717	·9649	·9575
4·60	·9989	·9967	·9937	·9899	·9855	·9804	·9747	·9686	·9620
4·65	·9990	·9971	·9944	·9911	·9871	·9825	·9775	·9719	·9660
4·70	·9991	·9974	·9951	·9921	·9885	·9845	·9799	·9750	·9696
4·75	0·9992	0·9977	0·9956	0·9930	0·9898	0·9862	0·9822	0·9777	0·9729
4·80	·9993	·9980	·9962	·9938	·9910	·9878	·9842	·9802	·9759
4·85	·9994	·9983	·9966	·9945	·9920	·9892	·9860	·9824	·9786
4·90	·9995	·9985	·9970	·9952	·9930	·9904	·9876	·9844	·9810
4·95	·9995	·9987	·9974	·9958	·9938	·9916	·9890	·9862	·9832
5·00	0·9996	0·9988	0·9977	0·9963	0·9946	0·9926	0·9903	0·9878	0·9851

Table 1 (cont.). *Probability integral of the range W in normal samples of size n*

W \ n	11	12	13	14	15	16	17	18	19	20
2·50	0·2007	0·1644	0·1342	0·1094	0·0890	0·0722	0·0586	0·0474	0·0383	0·0309
2·55	·2213	·1833	·1514	·1247	·1026	·0842	·0690	·0565	·0462	·0377
2·60	·2429	·2033	·1697	·1413	·1174	·0974	·0807	·0668	·0552	·0455
2·65	·2653	·2243	·1891	·1591	·1336	·1120	·0937	·0783	·0654	·0545
2·70	·2885	·2462	·2096	·1780	·1509	·1278	·1080	·0911	·0768	·0647
2·75	0·3124	0·2690	0·2311	0·1981	0·1696	0·1449	0·1236	0·1053	0·0896	0·0761
2·80	·3368	·2926	·2536	·2194	·1894	·1632	·1405	·1208	·1037	·0889
2·85	·3617	·3169	·2770	·2416	·2103	·1829	·1587	·1376	·1192	·1031
2·90	·3870	·3417	·3011	·2647	·2324	·2036	·1782	·1558	·1360	·1186
2·95	·4126	·3670	·3258	·2887	·2554	·2255	·1989	·1752	·1542	·1355
3·00	0·4382	0·3927	0·3512	0·3134	0·2792	0·2484	0·2207	0·1959	0·1737	0·1538
3·05	·4639	·4186	·3769	·3387	·3039	·2723	·2436	·2178	·1944	·1734
3·10	·4895	·4446	·4029	·3645	·3292	·2970	·2675	·2407	·2164	·1943
3·15	·5150	·4706	·4292	·3907	·3551	·3224	·2923	·2647	·2394	·2164
3·20	·5401	·4965	·4555	·4171	·3814	·3483	·3177	·2895	·2635	·2396
3·25	0·5649	0·5222	0·4817	0·4437	0·4081	0·3748	0·3438	0·3151	0·2885	0·2638
3·30	·5893	·5475	·5078	·4703	·4348	·4016	·3704	·3413	·3142	·2890
3·35	·6131	·5725	·5337	·4967	·4617	·4286	·3974	·3681	·3407	·3150
3·40	·6363	·5970	·5592	·5230	·4885	·4557	·4246	·3953	·3677	·3417
3·45	·6589	·6209	·5842	·5489	·5151	·4827	·4519	·4227	·3950	·3689
3·50	0·6807	0·6442	0·6087	0·5744	0·5413	0·5096	0·4792	0·4502	0·4226	0·3964
3·55	·7017	·6668	·6326	·5994	·5672	·5362	·5063	·4777	·4504	·4242
3·60	·7220	·6886	·6558	·6237	·5926	·5624	·5332	·5051	·4781	·4522
3·65	·7414	·7096	·6782	·6474	·6173	·5881	·5596	·5321	·5056	·4801
3·70	·7600	·7298	·6998	·6704	·6414	·6132	·5856	·5588	·5329	·5078
3·75	0·7776	0·7491	0·7206	0·6925	0·6648	0·6376	0·6110	0·5850	0·5598	0·5352
3·80	·7944	·7675	·7406	·7138	·6873	·6613	·6357	·6106	·5861	·5622
3·85	·8103	·7850	·7596	·7342	·7090	·6841	·6596	·6355	·6118	·5887
3·90	·8254	·8016	·7777	·7537	·7298	·7061	·6827	·6596	·6369	·6145
3·95	·8395	·8173	·7948	·7723	·7497	·7273	·7050	·6829	·6611	·6397
4·00	0·8528	0·8321	0·8111	0·7899	0·7686	0·7474	0·7263	0·7053	0·6845	0·6640
4·05	·8653	·8460	·8264	·8065	·7866	·7666	·7466	·7268	·7070	·6874
4·10	·8769	·8590	·8408	·8223	·8036	·7848	·7660	·7472	·7285	·7099
4·15	·8878	·8712	·8543	·8371	·8196	·8021	·7844	·7667	·7491	·7315
4·20	·8978	·8826	·8669	·8509	·8347	·8183	·8018	·7852	·7686	·7520
4·25	0·9072	0·8931	0·8787	0·8639	0·8488	0·8336	0·8182	0·8027	0·7871	0·7715
4·30	·9159	·9029	·8896	·8760	·8620	·8479	·8336	·8191	·8046	·7899
4·35	·9238	·9120	·8998	·8872	·8744	·8613	·8480	·8345	·8210	·8073
4·40	·9312	·9204	·9092	·8976	·8858	·8737	·8614	·8490	·8364	·8237
4·45	·9379	·9281	·9178	·9073	·8964	·8853	·8740	·8625	·8508	·8391
4·50	0·9441	0·9352	0·9258	0·9162	0·9062	0·8960	0·8856	0·8750	0·8643	0·8534
4·55	·9498	·9417	·9332	·9244	·9153	·9060	·8964	·8867	·8768	·8667
4·60	·9550	·9476	·9399	·9319	·9236	·9151	·9064	·8975	·8884	·8791
4·65	·9597	·9530	·9460	·9388	·9313	·9235	·9155	·9074	·8991	·8906
4·70	·9640	·9579	·9516	·9451	·9383	·9312	·9240	·9165	·9090	·9012
4·75	0·9678	0·9624	0·9567	0·9508	0·9446	0·9383	0·9317	0·9249	0·9180	0·9110
4·80	·9713	·9665	·9614	·9560	·9505	·9447	·9387	·9326	·9264	·9199
4·85	·9745	·9702	·9656	·9608	·9558	·9505	·9452	·9396	·9340	·9281
4·90	·9774	·9735	·9694	·9650	·9605	·9559	·9510	·9460	·9409	·9356
4·95	·9799	·9765	·9728	·9689	·9649	·9607	·9563	·9518	·9472	·9424
5·00	0·9822	0·9791	0·9759	0·9724	0·9688	0·9650	0·9611	0·9571	0·9529	0·9486

Table 1 (cont.). *Probability integral of the range W in normal samples of size n*

W \ n	2	3	4	5	6	7	8	9	10
5·00	0·9996	0·9988	0·9977	0·9963	0·9946	0·9926	0·9903	0·9878	0·9851
5·05	·9996	·9990	·9980	·9967	·9952	·9935	·9915	·9893	·9869
5·10	·9997	·9991	·9982	·9971	·9958	·9942	·9925	·9906	·9884
5·15	·9997	·9992	·9985	·9975	·9963	·9950	·9934	·9917	·9898
5·20	·9998	·9993	·9986	·9978	·9968	·9956	·9942	·9927	·9911
5·25	0·9998	0·9994	0·9988	0·9981	0·9972	0·9961	0·9949	0·9936	0·9922
5·30	·9998	·9995	·9990	·9983	·9975	·9966	·9956	·9944	·9931
5·35	·9998	·9995	·9991	·9985	·9979	·9971	·9961	·9951	·9940
5·40	·9999	·9996	·9992	·9987	·9981	·9974	·9966	·9957	·9948
5·45	·9999	·9997	·9993	·9989	·9984	·9978	·9971	·9963	·9954
5·50	0·9999	0·9997	0·9994	0·9991	0·9986	0·9981	0·9975	0·9968	0·9960
5·55	·9999	·9997	·9995	·9992	·9988	·9983	·9978	·9972	·9965
5·60	·9999	·9998	·9996	·9993	·9989	·9985	·9981	·9976	·9970
5·65	·9999	·9998	·9996	·9994	·9991	·9987	·9983	·9979	·9974
5·70	0·9999	·9998	·9997	·9995	·9992	·9989	·9986	·9982	·9977
5·75	1·0000	0·9999	0·9997	0·9995	0·9993	0·9991	0·9988	0·9984	0·9981
5·80		·9999	·9998	·9996	·9994	·9992	·9989	·9986	·9983
5·85		·9999	·9998	·9997	·9995	·9993	·9991	·9988	·9986
5·90		·9999	·9998	·9997	·9996	·9994	·9992	·9990	·9988
5·95		·9999	·9998	·9998	·9996	·9995	·9993	·9991	·9989
6·00		0·9999	0·9999	0·9998	0·9997	0·9996	0·9994	0·9993	0·9991
6·05		·9999	·9999	·9998	·9997	·9996	·9995	·9994	·9992
6·10		0·9999	·9999	·9998	·9998	·9997	·9996	·9995	·9993
6·15		1·0000	·9999	·9999	·9998	·9997	·9996	·9995	·9994
6·20			·9999	·9999	·9998	·9998	·9997	·9996	·9995
6·25			0·9999	0·9999	0·9999	0·9998	0·9997	0·9997	0·9996
6·30			0·9999	·9999	·9999	·9998	·9998	·9997	·9996
6·35			1·0000	·9999	·9999	·9999	·9998	·9998	·9997
6·40				0·9999	·9999	·9999	·9998	·9998	·9997
6·45				1·0000	·9999	·9999	·9999	·9998	·9998
6·50					0·9999	0·9999	0·9999	0·9999	0·9998
6·55					0·9999	·9999	·9999	·9999	·9998
6·60					1·0000	·9999	·9999	·9999	·9999
6·65						0·9999	·9999	·9999	·9999
6·70						1·0000	0·9999	·9999	·9999
6·75							1·0000	0·9999	0·9999
6·80								0·9999	·9999
6·85								1·0000	·9999
6·90									0·9999
6·95									1·0000
7·00									
7·05									
7·10									
7·15									
7·20									
7·25									

Table 1 (cont.). *Probability integral of the range W in normal samples of size n*

W \ n	11	12	13	14	15	16	17	18	19	20
5·00	0·9822	0·9791	0·9759	0·9724	0·9688	0·9650	0·9611	0·9571	0·9529	0·9486
5·05	·9843	·9815	·9786	·9756	·9723	·9690	·9655	·9618	·9581	·9543
5·10	·9861	·9837	·9811	·9784	·9755	·9725	·9694	·9661	·9628	·9593
5·15	·9878	·9856	·9833	·9809	·9783	·9757	·9729	·9700	·9670	·9639
5·20	·9893	·9874	·9853	·9832	·9809	·9785	·9760	·9735	·9708	·9681
5·25	0·9906	0·9889	0·9871	0·9852	0·9832	0·9811	0·9789	0·9766	0·9742	0·9718
5·30	·9917	·9903	·9887	·9870	·9852	·9833	·9814	·9794	·9773	·9751
5·35	·9928	·9915	·9901	·9886	·9870	·9854	·9836	·9819	·9800	·9781
5·40	·9937	·9925	·9913	·9900	·9886	·9872	·9856	·9841	·9824	·9807
5·45	·9945	·9935	·9924	·9912	·9900	·9888	·9874	·9860	·9846	·9831
5·50	0·9952	0·9943	0·9934	0·9924	0·9913	0·9902	0·9890	0·9878	0·9865	0·9852
5·55	·9958	·9951	·9942	·9933	·9924	·9914	·9904	·9893	·9882	·9870
5·60	·9964	·9957	·9950	·9942	·9934	·9925	·9916	·9907	·9897	·9887
5·65	·9969	·9963	·9956	·9950	·9943	·9935	·9927	·9919	·9910	·9901
5·70	·9973	·9968	·9962	·9956	·9950	·9944	·9937	·9929	·9922	·9914
5·75	0·9976	0·9972	0·9967	0·9962	0·9957	0·9951	0·9945	0·9939	0·9932	0·9925
5·80	·9980	·9976	·9972	·9967	·9963	·9958	·9952	·9947	·9941	·9935
5·85	·9982	·9979	·9976	·9972	·9968	·9963	·9959	·9954	·9949	·9944
5·90	·9985	·9982	·9979	·9976	·9972	·9968	·9964	·9960	·9956	·9952
5·95	·9987	·9985	·9982	·9979	·9976	·9973	·9969	·9966	·9962	·9958
6·00	0·9989	0·9987	0·9984	0·9982	0·9979	0·9977	0·9974	0·9971	0·9967	0·9964
6·05	·9990	·9989	·9987	·9984	·9982	·9980	·9977	·9975	·9972	·9969
6·10	·9992	·9990	·9989	·9987	·9985	·9983	·9981	·9978	·9976	·9973
6·15	·9993	·9992	·9990	·9989	·9987	·9985	·9983	·9981	·9979	·9977
6·20	·9994	·9993	·9992	·9990	·9989	·9987	·9986	·9984	·9982	·9980
6·25	0·9995	0·9994	0·9993	0·9992	0·9991	0·9989	0·9988	0·9986	0·9985	0·9983
6·30	·9996	·9995	·9994	·9993	·9992	·9991	·9990	·9988	·9987	·9986
6·35	·9996	·9996	·9995	·9994	·9993	·9992	·9991	·9990	·9989	·9988
6·40	·9997	·9996	·9996	·9995	·9994	·9993	·9992	·9992	·9991	·9990
6·45	·9997	·9997	·9996	·9996	·9995	·9994	·9994	·9993	·9992	·9991
6·50	0·9998	0·9997	0·9997	0·9996	0·9996	0·9995	0·9995	0·9994	0·9993	0·9993
6·55	·9998	·9998	·9997	·9997	·9996	·9996	·9995	·9995	·9994	·9994
6·60	·9998	·9998	·9998	·9997	·9997	·9997	·9996	·9996	·9995	·9995
6·65	·9999	·9998	·9998	·9998	·9997	·9997	·9997	·9996	·9996	·9995
6·70	·9999	·9999	·9998	·9998	·9998	·9998	·9997	·9997	·9997	·9996
6·75	0·9999	0·9999	0·9999	0·9999	0·9998	0·9998	0·9998	0·9997	0·9997	0·9997
6·80	·9999	·9999	·9999	·9999	·9998	·9998	·9998	·9998	·9998	·9997
6·85	·9999	·9999	·9999	·9999	·9999	·9999	·9998	·9998	·9998	·9998
6·90	0·9999	·9999	·9999	·9999	·9999	·9999	·9999	·9998	·9998	·9998
6·95	1·0000	0·9999	·9999	·9999	·9999	·9999	·9999	·9999	·9999	·9998
7·00		1·0000	0·9999	0·9999	0·9999	0·9999	0·9999	0·9999	0·9999	0·9999
7·05			0·9999	0·9999	·9999	·9999	·9999	·9999	·9999	·9999
7·10			1·0000	1·0000	0·9999	0·9999	·9999	·9999	·9999	·9999
7·15					1·0000	1·0000	0·9999	0·9999	·9999	·9999
7·20							1·0000	1·0000	0·9999	0·9999
7·25									1·0000	1·0000

3. *The origin of the present tables*

Tables giving the expected or mean value and the standard deviation of range in random samples from the normal population of equation (1) were calculated by L. H. C. Tippett (1925) in the Department of Applied Statistics, University College, London. Since the probability distribution $f_n(w)$ is itself far from normal in form, it was evident that its mean and standard deviation alone would not provide all the information generally needed in practice. Tippett included in his paper some values of the constants β_1 and β_2 of the distribution and his work was extended by the present writer (Pearson, 1926, 1932) who

Table 2

Size of sample n	Factor a_n	Lower percentage points						Upper percentage points					
		0·1	0·5	1·0	2·5	5·0	10·0	10·0	5·0	2·5	1·0	0·5	0·1
2	0·8862	0·00	0·01	0·02	0·04	0·09	0·18	2·33	2·77	3·17	3·64	3·97	4·65
3	0·5908	0·06	0·13	0·19	0·30	0·43	0·62	2·90	3·31	3·68	4·12	4·42	5·06
4	0·4857	0·20	0·34	0·43	0·59	0·76	0·98	3·24	3·63	3·98	4·40	4·69	5·31
5	0·4299	0·37	0·55	0·66	0·85	1·03	1·26	3·48	3·86	4·20	4·60	4·89	5·48
6	0·3946	0·54	0·75	0·87	1·06	1·25	1·49	3·66	4·03	4·36	4·76	5·03	5·62
7	0·3698	0·69	0·92	1·05	1·25	1·44	1·68	3·81	4·17	4·49	4·88	5·15	5·73
8	0·3512	0·83	1·08	1·20	1·41	1·60	1·83	3·93	4·29	4·61	4·99	5·26	5·82
9	0·3367	0·96	1·21	1·34	1·55	1·74	1·97	4·04	4·39	4·70	5·08	5·34	5·90
10	0·3249	1·08	1·33	1·47	1·67	1·86	2·09	4·13	4·47	4·79	5·16	5·42	5·97
11	0·3152	1·20	1·45	1·58	1·78	1·97	2·20	4·21	4·55	4·86	5·23	5·49	6·04
12	0·3069	1·30	1·55	1·68	1·88	2·07	2·30	4·29	4·62	4·92	5·29	5·54	6·09

Estimate of $\sigma = a_n \times$ range (or mean range) in a sample of n observations.

developed an approximate method of determining probability levels for w and provided some provisional tables of these. The need has, however, been felt for some time for a full and accurate table of the probability integral of the range to fit into place among other fundamental tables associated with the normal distribution. The completion of this objective has been made possible by a grant from the Department of Scientific and Industrial Research, whose assistance in the matter is acknowledged with warm appreciation. The actual method of computation was planned by Dr H. O. Hartley and the calculations were carried out under his supervision by Scientific Computing Service, Ltd. The scope of the main table was limited to $n \leqslant 20$. As n increases beyond this value there is an increasing risk that the table may be misleading in practice, since $f_n(w)$ becomes very sensitive to relatively slight departures from normality in the tails of the population distribution.

REFERENCES

DAVIES, O. L. & PEARSON, E. S. (1934). *J.R. Statist. Soc.* Supp. **1**, 76.

DUDDING, B. P. & JENNETT, W. J. (1942). *The Application of Statistical Methods to Quality Control.* British Standards Institution, No. 600, (1942).

McKAY, A. T. & PEARSON, E. S. (1933). *Biometrika*, **25**, 415.

PEARSON, E. S. (1926). *Biometrika*, **18**, 173.

—— (1932). *Biometrika*, **24**, 404.

—— (1935). *The Application of Statistical Methods to Industrial Standardization and Quality Control.* British Standards Institution, No. 600, (1935).

TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364.

II. NUMERICAL EVALUATION OF THE PROBABILITY INTEGRAL

By H. O. HARTLEY

The formula used for the tabulation of the probability integral $P_n(W)$ of the range in normal samples of size n is given in the paper printed on pp. 334–48 below, where it proved that

$$P_n(W) = \left(\int_{-\frac{1}{2}W}^{+\frac{1}{2}W} z(x)\,dx\right)^n + 2n\int_{\frac{1}{2}W}^{\infty} z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^{n-1} du, \tag{1}$$

where $$z(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}.$$

Certain properties of this formula and the facilities provided by certain modern calculating machines make this integral amenable to tabulation.

The main work consists in the evaluation of the integral

$$2n\int_{\frac{1}{2}W}^{\infty} z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^{n-1} du, \tag{2}$$

by quadrature for a two variable network of values of n and W. The range of integration is from $\frac{1}{2}W$ up to a point where the integrand

$$z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^{n-1} \tag{3}$$

vanishes to 7-decimal accuracy.* For each point of the network n, W, therefore, the integrand (3) was tabulated for a set of equidistant values of u covering the range of integration. The interval of integration was chosen as $\Delta u = 0{\cdot}2$ throughout. This was sufficient to obtain about 5-decimal accuracy in the integral (2).

The interval in W was taken as wide as possible but sufficiently fine to permit checking by differencing and the subsequent subtabulation of $P_n(W)$ to interval 0·05, which is the interval in the final table. An interval of $\Delta W = 0{\cdot}25$ was therefore chosen for the n, W network.

For small values of W it was necessary to tabulate the integrand for all integers n for which $P_n(W)$ is required in the final table. For larger values of n and W, however, it was sufficient to calculate the final integral (1) for odd n and to obtain intermediate values by interpolation. Below, then, is shown the two variable network for which the integral (2) was produced by quadrature:

$$\left.\begin{array}{llll} W = 0{\cdot}00\ (0{\cdot}25)\ 1{\cdot}25 & \text{and} & n = 3\ (1)\ 20. \\ W = 1{\cdot}50\ (0{\cdot}25)\ 2{\cdot}75 & & n = 3\ (1)\ \ 9\ (2)\ 23. \\ W = 3{\cdot}00\ (0{\cdot}25)\ 3{\cdot}25 & & n = 3\ (1)\ \ 5\ (2)\ 23. \\ W = 3{\cdot}50\ (0{\cdot}25)\ 8{\cdot}00 & & n = 3\ (2)\ 23. \end{array}\right\} \tag{4}$$

For $n = 2$ the final integral $P_2(W)$ is given directly by the normal integral and may be obtained by interpolation in Table II of *Tables for Statisticians and Biometricians*, Part I. Using the notation of that table (Sheppard's original notation) we have

$$P_2(W) = \alpha\left(\frac{W}{\sqrt{2}}\right).$$

Moreover, for purposes of interpolation, use was made of the formal relation

$$P_1(W) = 1 \quad \text{for} \quad W > 0.$$

For fixed u and W and for values of n in the arithmetic progression (4), the integrand (3) is a geometric progression with

$$z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^2$$

* The integrand was calculated to 7-decimal accuracy in order to obtain $P_n(W)$ to about 5-decimal accuracy.

as leading term and $\left(\int_{u-W}^{u} z(x)\,dx\right)$ or $\left(\int_{u-W}^{u} z(x)\,dx\right)^2$

as common ratio. This leading term as well as the common ratios were easily obtained from Table II of *Tables for Statisticians and Biometricians*, Part I and the terms of the progression were then automatically produced on a Mercedes calculating machine Model 38 M.S. and copied down in two-way tables with u as row heading, n as column heading and W as table heading. The values of the integrand were then checked by differencing column-wise and added to yield the main term of the integral (2). The correction terms which, according to Gregory's formula, convert the integrand-sum into the integral were calculated from the differences and checked by the application of Gauss' formula of integration. Finally, to obtain $P_n(W)$ the term

$$\left(\int_{-\frac{1}{2}W}^{+\frac{1}{2}W} z(x)\,dx\right)^n$$

was produced by continued multiplication on the Mercedes and added to the corresponding integral (2) to yield $P_n(W)$ for all points of the above network.

For odd values of n the integral $P_n(W)$ was then differenced W-wise on the National machine which, incidentally, produced column totals $\sum_W P_n(W)$ for these values of n. Two checks were applied at this stage. One consisted in inspecting the fourth order differences. As a second check, the mean range, $\bar{w}_n$, was calculated from the formula

$$\bar{w}_n = 8 - \int_0^8 P_n(w)\,dw,*$$

and compared with the correct mean range given in Table XXII of *Tables for Statisticians and Biometricians*, Part II. Finally, the function $P_n(W)$ was subtabulated to interval 0·05 on the National machine by a method similar to that described in detail by L. J. Comrie (1936).

The values of $P_n(W)$ for even n were then obtained by interpolation with the help of two interpolation formulae of Lagrangian type:

$$\begin{aligned} 2048P_n(W) = &-5[P_{n-7}(W)+P_{n+7}(W)]+49[P_{n-5}(W)+P_{n+5}(W)] \\ &-245[P_{n-3}(W)+P_{n+3}(W)]+1225[P_{n-1}(W)+P_{n+1}(W)], \end{aligned} \tag{5}$$

$$\begin{aligned} 20P_n(W) = &P_{n-3}(W)+P_{n+3}(W)-6[P_{n-2}(W)+P_{n+2}(W)] \\ &+15[P_{n-1}(W)+P_{n+1}(W)]. \end{aligned} \tag{6}$$

Formula (5) yields the interpolate for even n from the given values of $P_n(W)$ at adjacent odd values of n. This formula was used throughout. In some cases, however, the resulting interpolate was accurate to about 3 places of decimals only. In such cases values of P_{n-3}, P_{n+3}, P_{n-1}, P_{n+1} accurate to 5 places of decimals and values of P_{n-2}, P_{n+2} accurate to (say) 3 places of decimals were substituted in formula (6). This yielded a 'corrected value' of $P_n(W)$. The process was then repeated for $n = n+2$ and so on until all values of $P_n(W)$ had 'settled down' for even values of n. It is easy to see that the process is convergent and that the maximum error in the interpolate is 2 units for the 5th decimal.

After completion of the interpolation n-wise, the interpolates $P_n(W)$ for even n were differenced W-wise, checked and subtabulated as for odd values of n.

* This is true provided $P_n(8)=1$ to 6-decimal place accuracy.

REFERENCE

Comrie, L. J. (1936). *J.R. Statist. Soc.* Suppl. **3**, 87–114.

NOTES ON TESTING STATISTICAL HYPOTHESES

BY E. S. PEARSON

1. In July 1939, a few weeks before the opening of the present war, a Conference on the Application of the Calculus of Probabilities was held at Geneva under the auspices of the International Institute of Intellectual Co-operation (League of Nations). At the public session at which a paper by Prof. J. Neyman was presented and also subsequently in some informal discussions, a number of questions were raised:

(*a*) In choosing a test for a statistical hypothesis, is it possible or even necessary to specify the hypotheses alternative to that tested? Why should not a test be made to depend only on the form of law associated with the hypothesis tested? For example, Newton's hypothesis of gravitation was formulated and tested without any need to define alternative laws.

(*b*) Is the method of approach to these problems advocated by Prof. Neyman and myself applicable to testing the appropriateness of probability *laws* or only to testing hypotheses regarding the numerical values of *constants* contained in these laws?

After the conclusion of the conference, I set down some Notes for a few of the statisticians who had taken part in the discussions, hoping that at leisure they might feel stimulated to define their views on the subject more precisely. But almost before the Notes were despatched, war in Europe had intervened. The only reply which I received was from Prof. Gumbel, and this, after some unavoidable delay, has now taken shape in the contribution printed on pp. 317–33 below. In publishing this, it seems useful to add my own Notes, which are given with only minor verbal alterations in the following pages. They are in part a restatement with rather different emphasis of views expressed in a paper published four years ago (Pearson, 1938).

2. With regard to one of the points raised under (*a*) above, it should be remembered that a statistical hypothesis as defined by Neyman and myself is a hypothesis concerning the probability law of random variables. The gravitational hypothesis of Newton is not a statistical hypothesis in the sense defined; statistical methods may be introduced to test the Newtonian hypothesis, however, and they will involve tests of statistical hypotheses or 'significance tests' because it will be assumed that errors of observation exist which may be regarded as random variables, probably taken to follow the normal distribution law.

For example, on the Newtonian hypothesis, the angular co-ordinates of a planet measured from the earth as origin may at certain moments be given as $\xi = \xi_t$, $\eta = \eta_t$ ($t = 1, 2, \ldots$). If we have a number of observations of position x_t, y_t,

subject to observational error, the statistical problem will be to test whether these are consistent with the hypothetical position values ξ_t, η_t, or whether they suggest that ξ, η have some other different values at the moments of observation. Thus the 'alternatives' that we have immediately in mind will be alternative values for ξ, η, not alternative gravitational hypotheses. If, however, some alternative law of motion were proposed, so that we could specify definite values ξ_t', η_t' alternative to the values ξ_t, η_t of the Newtonian law, then undoubtedly we could choose a statistical test which would be particularly efficient in discriminating between the two alternatives. Such a course became possible when the Einstein hypothesis was formulated and the orbit of Mercury considered. But the absence of an alternative gravitational law does not prevent us selecting a statistical test which will be (*a*) sensitive to departures in ξ, η from ξ_t, η_t, but (*b*) relatively insensitive to departures from normality in the distribution of errors. We should make this selection because, if the Newtonian law were incorrect, we believe that this would result in a change in ξ_t, η_t but not in a departure of the distribution of observational errors from the normal law.

This example, of course, concerns a statistical hypothesis regarding the values of two parameters ξ, η, not regarding the form of a probability law of random variables. The following general approach shows, however, that the principles discussed may be applied to testing hypotheses regarding probability laws.

3. Suppose that x is a continuous random variable and that H_0 is a statistical hypothesis which assumes that the elementary probability law for x is $p(x \mid H_0)$ in the interval $-\infty$ to $+\infty$. Thus

$$\int_{-\infty}^{+\infty} p(x \mid H_0)\, dx = 1. \tag{1}$$

Now write
$$y = \int_{-\infty}^{x} p(x \mid H_0)\, dx. \tag{2}$$

y will be a non-decreasing function of x having values confined to the interval $(0, 1)$. Further, the elementary probability law of y will be

$$p(y) = p(x) \Big/ \frac{dy}{dx} = 1 \quad \text{for} \quad 0 \leqslant y \leqslant 1, \tag{3}$$

or all values of y between 0 and 1 are 'equally probable'.

Suppose now that we wish to use a set of n independent values $x_1, x_2, \ldots, x_n$ to test that the probability law is of the assumed form $p(x \mid H_0)$. It is clear that the hypothesis H_0 is exactly equivalent to the hypothesis, say h_0, that the n values $y_1, y_2, \ldots, y_n$ (obtained from the x's by the transformation (2)) have been sampled subject to the probability law (3). Just as the point $(x_1, x_2, \ldots, x_n)$ may be represented in an unlimited n-dimensioned space having probability density

$$p(x_1, x_2, \ldots, x_n \mid H_0) = \prod_{i=1}^{n} \{p(x_i \mid H_0)\}, \tag{4}$$

if H_0 is true, so the point $(y_0, y_1, \ldots, y_n)$ may be placed in an n-dimensioned hypercube with sides of unit length and with uniform probability density, if H_0 and therefore h_0 is true. It follows that if H_0 is what has been termed a 'simple hypothesis', i.e. specifies the form of $p(x \mid H_0)$ completely,* then the test of H_0 may always be transformed to the test of h_0. If then it were correct to say that the test of a statistical hypothesis *depends only on the form of the law specified by* H_0, it follows that for the type of situation considered the testing of a statistical hypothesis could always be reduced to the following simple problem:

To test whether a sample of n independent random variables $y_1, y_2, \ldots, y_n$ $(0 \leqslant y_i \leqslant 1)$ has been selected from the so-called rectangular distribution, i.e. the distribution for which $p(y) = 1$, $(0 \leqslant y \leqslant 1)$.

4. We are at once faced, therefore, with the question of how to test this simple but apparently fundamental hypothesis. If h_0 is true, the sample point is equally likely to fall at any point within the n-dimensioned hypercube. Thus in picking out the critical (or rejection) region in this space we can get no assistance whatsoever from the changes in probability density, as we might do in the x-space. If we wish to use a level of significance of α (say $\alpha = 0{\cdot}01$) for rejecting h_0, it is clear that an infinite number of critical regions satisfying this condition are available; it is only necessary to select a region whose content is α.

If we consider the n values of y and plot them in the interval (0, 1) as follows,

Fig. 1.

the great majority of samples, from a rectangular distribution, at any rate if n is not too small, will be spread out fairly uniformly throughout the interval. Perhaps an 'ideal' sample against which to measure irregularities might be described as one for which the values of y fell at

$$\frac{1}{2n},\ \frac{3}{2n},\ \frac{5}{2n},\ \ldots,\ \frac{2n-1}{2n}.$$

But what form of departure from this ideal of uniformity are we to pick out as suggesting that the hypothesis h_0 is disproved? Should we judge significance by paying attention to the value of the mean y, of the variance, of the range of variation or of higher moments? Or should we use the χ^2 or ω^2 tests? It seems difficult to find any basis for choice which could be regarded in any sense as the 'best'. For any set of values $y_1, y_2, \ldots, y_n$ *some* critical region of size α can always be found which will contain the sample point and therefore lead to the rejection of h_0. Indeed, the task of selecting a unique region on any rational basis would seem to be insoluble.

* This condition is important. If the values of certain constants contained in the probability law need to be estimated from the observations, then the n values of y will not form a true random sample from a rectangular distribution. They will be subject to certain limitations to their degrees of freedom, though these may be relatively unimportant if n is large.

5. Directly it is recognized, however, that the choice of a test of a statistical hypothesis depends on something more than the form of the law associated with that hypothesis, it can be seen how a solution may be obtained. If we can specify a single alternative H_1 to H_0 or a class of alternatives $C(H)$, then we shall have also an alternative h, or a class $C(h)$ to h_0. Thus, if $p(x \mid H_1)$ denotes a probability law alternative to $p(x \mid H_0)$, then for y the alternative is

$$p(y \mid h_1) = p(x \mid H_1) \Big/ \frac{dy}{dx} = \frac{p(x \mid H_1)}{p(x \mid H_0)}\bigg|_{x=f(y)} \quad \text{for} \quad 0 \leqslant y \leqslant 1, \tag{5}$$

where $f(y)$ means the solution of

$$y = \int_{-\infty}^{x} p(x \mid H_0)\, dx \tag{6}$$

with regard to x. For example, Fig. 2 shows three typical forms of alternative $p(y \mid h_1)$, $p(y \mid h_2)$ and $p(y \mid h_3)$ associated with alternatives $p(x \mid H_1)$, ..., etc., to $p(x \mid H_0)$.

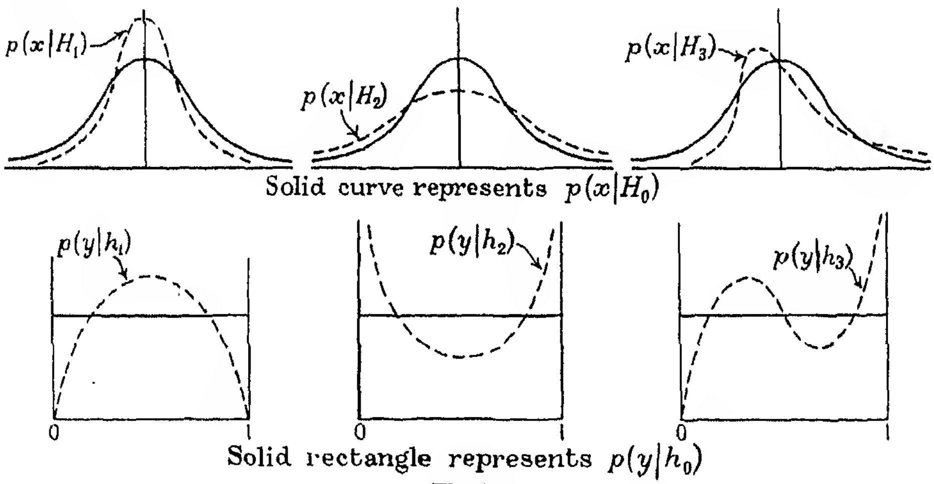


Fig. 2.

We can now see the kind of test which will be most efficient for testing H_0 with regard to possible classes of alternatives. If the alternative laws are of smaller dispersion (as $p(x \mid H_1)$), we must be on the look-out for too many values of y near $\frac{1}{2}$ and too few near 0 and 1. For alternatives with greater dispersion (as $p(x \mid H_2)$), we must reject H_0 when there are too many y's near 0 or 1 and too few near $\frac{1}{2}$. While if the alternatives are likely to be asymmetrical curves (as $p(x \mid H_3)$), then a different rule will be needed, as suggested by the $p(y \mid h_3)$ curve.

6. It follows that in so far as it is possible to formulate the class of admissible probability laws $p(x \mid H)$, the problem of selecting the most efficient test of H_0 reduces to that of choosing a critical region in the n-dimensioned hypercube which is most effective in detecting, from a sample of n values of y, differences between the rectangle $p(y \mid h_0)$ and the appropriate alternative forms $p(y \mid h)$.

If H_1 is a single admissible alternative, then it has been shown (Neyman & Pearson, 1933, p. 298) that the region w_0 of content α in the hypercube, within which

$$\frac{\prod_{i=1}^{n} p(y_i \mid h_0)}{\prod_{i=1}^{n} p(y_i \mid h_1)} < k, \tag{7}$$

or, in view of equation (3),

$$\prod_{i=1}^{n} p(y_i \mid h_1) > \frac{1}{k}, \tag{8}$$

where k is chosen so that

$$P\{(y_1, y_2, \ldots, y_n) \epsilon w_0 \mid h_0\} = \alpha \tag{9}$$

has the following property.

Of all regions of content α, w_0 is more likely than any other to include the sample point when h_1, and not h_0, is true. The region has been termed the best critical region for testing h_0 with regard to the alternative h_1.

As soon as H_0 and H_1 are specified, clearly $p(y \mid h_1)$ and therefore the region w_0 can be found, although mathematically it may be rather difficult to determine the appropriate boundary $\prod_{i=1}^{n} p(y_i \mid h_1) = \text{constant}$, so as to satisfy (8). Since this product is the probability density in the hypercube given by h_1, it will be seen that what we set out to do is to include in the critical region those parts of the sample space where the density for h_1 is highest. It is here, on repeated sampling, that sample points would tend to be concentrated if h_1 is true, instead of being uniformly distributed as under h_0.

7. If instead of a single alternative h_1, there is a class of admissible alternatives $C(h)$, there may or may not be common points of concentration that can be included in the critical region. This will depend on whether the inequality (8) above defines a region independent of the particular hypothesis h of the class $C(h)$. Even if there is no single region of content α which is exactly a 'best critical region' for h_0 with regard to all members of $C(h)$, the general principle may still be used as a guide. We build up a critical region out of those parts of the hypercube where the probability density tends to be concentrated when the probability law departs from $p(x \mid H_0)$ in the direction of the alternatives included in $C(H)$.

For example, in my earlier paper (Pearson, 1938) I suggested as appropriate in the following situation a test which, while not based on a common best critical region, was selected so as to include regions of greatest density associated with alternatives of $C(h)$. For the hypothesis tested,

$$p(x \mid H_0) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \tag{10}$$

The alternatives are asymmetrical curves with the same mean and standard deviation as (10). A typical alternative would be the Type III curve

$$p(x \mid H) = c(1 + \tfrac{1}{2}x\sqrt{\beta_1})^{\frac{4}{\beta_1}-1} e^{-\frac{2x}{\sqrt{\beta_1}}}, \tag{11}$$

whose form departs more and more from (10) as $\sqrt{\beta_1}$ increases from zero, but the class need not be defined as precisely as this. In this problem it appears that if n independent observations $x_1, x_2, \ldots, x_n$ are available, the following is a good test of H_0. Take as test function

$$Q = \prod_{i=1}^{n} (y'_i), \tag{12}$$

where

$$\left.\begin{aligned} y'_i &= 5(0\cdot2 - y_i) && \text{for} \quad 0 \leqslant y_i \leqslant 0\cdot2, \\ y'_i &= \tfrac{5}{3}(y_i - 0\cdot2) && \text{for} \quad 0\cdot2 < y_i \leqslant 0\cdot8, \\ y'_i &= 5(1 - y_i) && \text{for} \quad 0\cdot8 < y_i \leqslant 1, \end{aligned}\right\} \tag{13}$$

and

$$y_i = \int_{-\infty}^{x} \frac{1}{\sqrt{(2\pi)}}\, e^{-\frac{1}{2}x^2}\, dx. \tag{14}$$

If H_0 is true it may be shown that $-2\log_e Q$ is distributed as χ^2 with $2n$ degrees of freedom. Hence any desired significance level α, for Q, may be found. We should then reject H_0 when Q is significantly small.

A more systematic method of dealing with such problems has been considered by Neyman (1937) in his paper on 'smooth tests'.

8. To sum up, the position seems to be this. It has often been argued that a statistical test need only depend on the form of the probability law associated with the hypothesis tested. In the case where H_0 concerns the probability law of a single random variable and where $p(x \mid H_0)$ is precisely specified, by the transformation from x to y it has been shown that the problem of testing H_0 on the basis of n independent values of x can always be reduced to another problem, which involves this question. Can we regard a sample $y_1, y_2, \ldots, y_n$ as having been drawn from the rectangular distribution $p(y \mid h_0) = 1$, where $0 \leqslant y \leqslant 1$? We are faced with a single fundamental question and we have to consider whether it can be answered in a rational manner, unless we are prepared to take into account the kind of departures from the rectangular law that we either believe possible or at any rate consider it most important to be on the look out for.

The transformation from x to y seems to have the advantage that it concentrates attention on the main point at issue. That is my reason for emphasizing it in these Notes. Most of us have many preconceived ideas about appropriate tests if the probability law is taken in the form of $p(x \mid H_0)$; we are accustomed to use the mean, the standard deviation, certain functions of moments, the χ^2 test, ... But we are not so accustomed to test whether a sample comes from a rectangular distribution and we are therefore forced or, indeed, more willing to reconsider from first principles what course we should follow and why.

REFERENCES

Neyman, J. (1937). *Skand. Actuar. Tidskr.* 149.
Neyman, J. & Pearson, E. S. (1933). *Philos. Trans.* A, **231**, 289.
Pearson, E. S. (1938). *Biometrika*, **30**, 134.

SIMPLE TESTS FOR GIVEN HYPOTHESES

By E. J. GUMBEL
New School for Social Research, New York

CONTENTS

In dealing with the problem of testing statistical hypotheses J. Neyman (1937) and E. S. Pearson (1938) have considered the use of the probability integral transformation, which leads to a theoretical uniform distribution. This method presupposes that the usual comparison between theory and observations has already been applied. We shall first improve this comparison by introducing control curves. Then we shall apply to the uniform distribution the usual methods and the control curves. This will lead to simple tests for given statistical hypotheses.

1. Control curves and the probability integral transformation

Let x be a continuous random variable for which n observations have been made. Let x_m be the observed values arranged in increasing order of magnitude, m $(1, 2, \ldots, n)$ being the serial number. The simplest way of representing the observations is to plot the cumulative histogram x_m, m. The relative number $W^{(0)}(x_m)$ of observations less than or equal to x_m is given by

$$m = nW^{(0)}(x_m).$$

The consecutive differences

$$nW^{(0)}(x_m) - nW^{(0)}(x_l) \quad (l < m)$$

constitute the observed distribution. Many statisticians present, instead of the original observations x_m, only the number of cases within certain arbitrary classes. From the practical standpoint this means a simplification, from the theoretical standpoint a complication. We shall suppose that all x_m are known.

The choice of a probability density to be applied to the observations constitutes the hypothesis. The probability density $w(x, c_1, c_2, \ldots)$, where c_1, c_2, ... are the constants, is called the theoretical distribution. For sake of simplicity it is assumed that all observed values x_m have the same theoretical distribution. The probability $W(x, c_1, c_2, \ldots)$ of a value equal to or less than x is given by

$$W(x, c_1, c_2, \ldots) = \int_{-\infty}^{x} w(z, c_1, c_2, \ldots)\, dz, \tag{1}$$

where z is the variable of integration. It is customary to compare the observations x_m, m with the cumulative frequency curve x, $nW(x)$. This comparison can be improved in the following way: The mth observation is a statistical variable distributed according to

$$\mathfrak{w}_m(x, n) = \binom{n}{m} m W(x)^{m-1}(1 - W(x))^{n-m} w(x). \tag{2}$$

In a previous article (Gumbel, 1935) it has been shown that, for an ordinary unlimited distribution, with large n and with m of the size $\frac{1}{2}n$, the distribution (2) converges towards a normal distribution with a mean given by $W(x) = m/n$ and a standard deviation

$$\sigma = \frac{1}{w(x)} \sqrt{\frac{W(x)(1 - W(x))}{n}}. \tag{3}$$

This formula does not contain m explicitly. Since each theoretical value x can be interpreted as an mth value, (3) gives its standard deviation. The interval $x \mp \sigma$ will be called the *control interval*. Under the above condition, the probability that an mth value will fall within the control interval is about $\frac{2}{3}$. The two curves obtained by plotting $x \mp \sigma$, $nW(x)$ will be called control curves.

For a given initial distribution we shall have to find the mean and the standard deviation of the mth value which may differ from those of the general solution, especially if n is small. The control interval will be a certain function of $w(x)$. Also the probability associated with the interval $x \mp \sigma$ may differ from that of the general solution and may depend upon x. For the exponential distribution (Gumbel, 1937) the precision diminishes with increasing values of the variable. Below we shall apply this control to the uniform distribution $w(x) = \text{constant}$.

The calculation of the probability $W(x)$ and the control curves can often be simplified by an indirect method. For certain, but not all, distributions it is possible to eliminate the constants by introducing a new variable y as a function of x, where

$$x = f(y, c_1, c_2, \ldots). \tag{4}$$

Accordingly, the probability that a value of the transformed variable will fall in the interval y to $y + dy$ is

$$w(x)\,dx = w[f(y)]\,|f'(y)|\,dy.$$

We call

$$p(y) = w[f(y)]\,|f'(y)|,$$

or

$$p(y) = w(x)\left|\frac{dx}{dy}\right|, \tag{5}$$

the distribution of y. The probability $V(y)$ of a value equal to or less than y is

$$V(y) = W(x, c_1, c_2, \ldots). \tag{6}$$

The transformation (4) is chosen such that the expression $V(y)$ does not contain any constants which depend upon the observations. Therefore $V(y)$ can be

calculated once and for all as a function of y. Such tables have been calculated for the distributions for which this reduction is possible.

In order to compare the cumulative histogram x_m, m with the cumulative frequency x, $nW(x)$, and to use the control curves, it is first necessary to compute the constants c_1, c_2, ... If the method of moments is used, the area between the cumulative histogram and the horizontal line $W = 1$, the arithmetic mean, is conserved. The value of the variable x which corresponds to a selected, numerical value of $W(x)$ is obtained from the transformation (4).

A special case of the transformation (4) which leads to a reduced distribution of astounding simplicity is the probability integral transformation due to Karl Pearson (1933) who introduced for y the probability function $W(x)$. Since

$$\frac{dx}{dW} = \frac{1}{w(x)}, \tag{7}$$

we obtain from (5)

$$p(W) \equiv 1. \tag{8}$$

This identity means that the distribution of the probability is constant. As $p(W)$ is the probability density of a probability, it is difficult to establish its philosophical meaning. But formally the construction is valid and the corresponding value can be observed. It is our purpose to give several methods of judging the significance of the differences between the theoretical distribution (8) and the corresponding 'observed' distribution.

The word 'observation' will be given a special meaning. A certain theory h_0 which involves the choice of certain constants c_1, c_2, ... applied to the observations, leads to the values

$$W_0 = W(x_m, c_1, c_2, \dots \mid h_0) \quad (m = 1, 2, \dots, n),$$

corresponding to x_m. These values, contained in the interval 0, 1, are the 'observations'. To any other set of constants c_1', c_2', ... will correspond other 'observed' values

$$W_0' = W(x_m, c_1', c_2', \dots \mid h_0).$$

Therefore any test applied to the 'observations' of formula (8) might be used to judge the choice of the constants. To another hypothesis h_1 containing the constants d_1, d_2, ..., will correspond another set of 'observed' points

$$W_1 = W(x_m, d_1, d_2, \dots \mid h_1).$$

The same observations when interpreted by different theories or different constants lead to different 'observations'. An incorrect theory involving properly chosen constants might give better results than a correct theory involving improperly chosen constants. Therefore, to compare different theories, the constants for each must be determined with the same precision. In practice this condition will not always be fulfilled. For the precision of the characteristics depends upon the distribution for which they are calculated (Gumbel, 1936). Therefore, the

same method of determining the constants might lead to different degrees of precision for different distributions, whereas different determinations of the constants might lead to approximately the same precision.

There is another point for caution: all tests derived from the probability integral transformation apply to the analytic form of the hypotheses and at the same time to the choice of the constants. A formula containing many constants may reproduce the observations more closely than a formula containing few constants, even though the constants in the first hypothesis have no meaning. Therefore we must limit our comparison to hypotheses containing the same number of constants. For any set of statistical observations in the ordinary sense, there will usually correspond a small number of tenable hypotheses. We shall suppose it is known what they are. For we do not try to find a formula for the sake of doing it, but to explain the observed facts. We will not go so far as Neyman (1937), who formulated all possible alternatives by a series of orthogonal functions.

In theory the points representing $W(x_m)$ are distributed uniformly in the interval 0, 1. This is true for any hypothesis, provided the variable is continuous. But in practice this will never occur. The 'observed' points corresponding to any given hypothesis will differ from the theoretical set, even if the hypothesis is a very good one. The differences between the 'observed' set of points resulting from h_0 and h_1 and the theoretical set allow the construction of tests which can be used to judge which of two given hypotheses is the better. But no statistical method gives an answer to the question whether or not a hypothesis is true.

After a hypothesis has been selected, the preliminary steps which have to be made before it is possible to use the probability integral transformation, are: first, the determination of the constants; secondly, the calculation of probabilities $W(x)$ for the values of x given by the transformation (4); and thirdly, the calculation of the probabilities $W(x_m)$ of the observed values. It is only after these three operations have been carried out that we obtain the 'observations' which are to be compared with the theory (8). Therefore, any test based on the probability integral transformation presupposes the usual comparison of the observed cumulative histogram with the frequency curve. In many cases this comparison, checked by the control curves, will indicate a clear superiority of one of the theories. If this is true, there is no necessity for a new test.

It would be interesting to investigate the best criterion for judging the significance of the differences between the 'observed' and the uniform distribution (8). But for practical purposes it is sufficient to know whether the differences for h_0 are smaller or larger than for h_1. First we shall establish rough measures of comparison; afterwards, more refined ones.

2. CLASSICAL TESTS APPLIED TO A UNIFORM DISTRIBUTION

The comparison of an observed distribution of a continuous variable with the theoretical distribution is reduced by the probability integral transformation to a comparison of 'observed' points with a uniform distribution. It seems logical to use first the classical methods which are here very simple, as no constants have to be determined. For a uniform distribution

$$p(y) = 1 \quad (0 \leq y \leq 1),$$

the arithmetic mean and the median are, respectively,

$$\bar{y} = \tilde{y} = \tfrac{1}{2}. \tag{9}$$

The mean error θ and the probable error ρ, defined as half of the difference between the two quartiles, are

$$\theta = \rho = \tfrac{1}{4}. \tag{10}$$

The kth moment about the origin is

$$M_k = \frac{1}{k+1},$$

which gives the recurrence relation

$$\frac{1}{M_{k+1}} - \frac{1}{M_k} = 1. \tag{11}$$

Since the distribution is symmetrical, the odd moments about the arithmetic mean vanish. Therefore

$$\beta_1 = 0.$$

The even moments are

$$\mu_{2k} = \int_0^1 (y-\tfrac{1}{2})^{2k}\,dy = 2\int_0^{\frac{1}{2}} z^{2k}dz \quad \text{or} \quad \mu_{2k} = \frac{1}{2^{2k}(2k+1)}. \tag{12}$$

Therefore the standard deviation, the coefficient of variation and the second beta are, respectively,

$$\sigma = \frac{1}{2\sqrt{3}}, \tag{13}$$

$$v = \frac{1}{\sqrt{3}}, \tag{14}$$

and

$$\beta_2 = \tfrac{9}{5}. \tag{15}$$

It is necessary now to calculate the 'observed' means, the measures of dispersion, and the relations between successive moments. To control the agreement between the theoretical uniform distribution and the 'observed' points we can still employ the standard error $\sigma_{\bar{y}}$ of the arithmetic mean. The general formula

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

becomes according to (13)
$$\sigma_{\bar{y}} = \frac{1}{2\sqrt{(3n)}}. \tag{16}$$

The standard error of the dispersion for n large is $\sigma(\sigma^2) = \sqrt{\frac{\mu_4 - \mu_2^2}{n}}$, which becomes, according to (12),
$$\sigma(\sigma^2) = \sqrt{\left\{\frac{1}{4^2 n}\left(\frac{1}{5} - \frac{1}{9}\right)\right\}} = \frac{1}{6\sqrt{(5n)}}. \tag{17}$$

It seems reasonable to employ these old-fashioned tests before the use of more sophisticated methods is resorted to. Only if they fail would it be necessary to consider more elaborate methods.

The n 'observed' points $W(x_m)$ represent the probabilities, obtained from the hypothesis h_0, of the given observations in the ordinary sense, x_m. We plot these points in the interval 0, 1 which is divided into k cells of equal length, where k is chosen in such a way that n/k is an integer. If n is a multiple of 10, we choose the cells (0·0, 0·1), (0·1, 0·2), ..., (0·9, 1·0).

The probability density of a point falling somewhere within the interval 0, 1 is constant. As the interval is of length 1, this density is 1. Therefore, the probability of a point lying with a given cell is $1/k$, and the expected number of points in each cell is n/k. The 'observed' number of points obtained through a hypothesis h_0 will be a_ν ($\nu = 1, 2, ..., k$). If we apply another hypothesis h_1 to the same observations or introduce other numerical values for the constants, the new set of 'observed' points will lead to values b_ν which, in general, will differ from a_ν.

The classical statistical method of treating this material is the χ^2 test. As the numbers a_ν, b_ν will differ from the expected number n/k, we can calculate for both hypotheses
$$\chi_0^2 = \frac{k}{n}\sum_{\nu=1}^{k}\left(a_\nu - \frac{n}{k}\right)^2, \quad \chi_1^2 = \frac{k}{n}\sum_{\nu=1}^{k}\left(b_\nu - \frac{n}{k}\right)^2. \tag{18}$$

The better hypothesis will have a lower value of χ^2 and a greater value P, where P denotes the probability of obtaining the 'observed' deviations from uniform distribution or larger ones. The probability P depends upon the number of cells chosen. Therefore, to compare two competing hypotheses, the same division must be used.

The application of the χ^2 test to the 'observations' $W(x_m)$ eliminates an arbitrary action which is a serious and well-known drawback of the χ^2 test, when applied to the original observations x_m. The expected contents of the classes depend upon the distribution. Therefore certain classes, as a rule the first and the last, must be chosen such that the expected number is not too small, otherwise χ^2 becomes very large. In our case, no cell differs from any other and no arbitrary combination of cells is needed. We can choose $k = n$. The mean number of points in each cell will then be one and
$$\chi_0^2 = \sum_{\nu=1}^{n}(a_\nu - 1)^2, \quad \chi_1^2 = \sum_{\nu=1}^{n}(b_\nu - 1)^2. \tag{18'}$$

This choice removes another drawback of the χ^2 method: different classifications used for the same observations lead to different shapes of the distribution and therefore to different values of χ^2. Here the classification is prescribed once and for all.

Another comparison between theory and 'observation' may be based on the fact that different sets of points a_ν and b_ν have different probabilities. The probability that a_ν points will fall within the cell ν $(= 1, 2, \ldots, k)$ is

$$\frac{n!}{a_1!\,a_2!\ldots a_k!}\left(\frac{1}{k}\right)^{a_1+a_2+\ldots+a_k},$$

where $$a_1+a_2+\ldots+a_k = n.$$

Since the factor $n!\,k^{-n}$ is constant, it is sufficient to investigate

$$\Pi = \frac{1}{\prod\limits_{\nu=1}^{k} a_\nu!}. \tag{19}$$

Of course $\Pi < P$, as the latter probability applies to the 'observed' deviation or larger ones. The statement 'The probability for a_ν points to be contained in the cell ν is proportional to Π' may be inverted according to Bayes's principle. Therefore, Π is proportional to the probability that the distribution of points is rectangular, i.e. that h_0 is a good hypothesis.

The question for which set $\tilde{a}_\nu$ the probability Π is maximum, is the starting-point of the classical relation between entropy and probability. For large n the most probable set of points is the one which has the same number of points in each cell, i.e.

$$\tilde{a}_1 = \tilde{a}_2 = \ldots = \tilde{a}_k = \frac{n}{k}. \tag{20}$$

Let us call $\Pi_{\text{max.}}$ the probability which corresponds to this distribution. The probability of the hypothesis h_0 will be greater, equal to, or less than the probability of h_1, if

$$\frac{\Pi_0}{\Pi_{\text{max.}}} \gtreqless \frac{\Pi_1}{\Pi_{\text{max.}}}. \tag{21}$$

The relative probability of both hypotheses will be Π_1/Π_0 or Π_0/Π_1, depending on whether

$$\Pi_0 \gtrless \Pi_1.$$

As these probabilities depend on the number of cells, the same division must be used to compare two competing hypotheses. We can choose $k = n$. Then $\Pi_{\text{max.}} = 1$ and we can use Π_0 as test.

The entropy test (21) is closely related to the χ^2 test (18). This classical relation can be obtained in the following simple way: if q is the constant probability of a point falling within a given cell, then for n observations the expected

number of points within a cell is nq. But the 'observed' numbers a_ν will differ from the expected number by ϵ_ν so that

$$a_\nu = nq + \epsilon_\nu,$$

where
$$\sum_{\nu=1}^{k} \epsilon_\nu = 0.$$

The quotient (21) becomes
$$\frac{\Pi_0}{\Pi_{\text{max.}}} = \prod_{\nu=1}^{k} \frac{(nq)!}{(nq+\epsilon_\nu)!}.$$

When n is large, each factor becomes, by application of Stirling's formula,

$$\frac{nq!}{(nq+\epsilon_\nu)!} = \left(\frac{e}{nq}\right)^{\epsilon_\nu} \exp\left[-(nq+\epsilon_\nu+\tfrac{1}{2})\ln\left(1+\frac{\epsilon_\nu}{nq}\right)\right].$$

Expansion of the logarithm leads to

$$\frac{nq!}{(nq+\epsilon_\nu)!} = \left(\frac{e}{nq}\right)^{\epsilon_\nu} \exp\left[-\epsilon_\nu - \frac{\epsilon_\nu}{2nq} - \frac{\epsilon_\nu^2}{2nq}\right].$$

According to the meaning of ϵ_ν we obtain

$$\frac{\Pi_0}{\Pi_{\text{max.}}} = \exp\left[-\frac{1}{2}\sum_{\nu=1}^{k}\left(\frac{a_\nu - nq}{nq}\right)^2\right],$$

whence
$$\frac{\Pi_0}{\Pi_{\text{max.}}} = e^{-\frac{1}{2}\chi^2}. \tag{22}$$

Therefore, when n is large, the entropy test becomes identical with the χ^2 test. This result was derived by Neyman & Pearson (1928), when they showed that the χ^2 test followed from their 'likelihood ratio' method of approach.

Neither test will give an answer, if the number of points a_ν assigned by h_0 to the cell ν is equal to the number of points b_λ assigned by h_1 to the cell λ, $a_\nu = b_\lambda$, where for any ν $(= 1, 2, \ldots, k)$ it is possible to find a λ $(= 1, 2, \ldots, k)$, such that not all $\lambda = \nu$. An example of this occurring is shown in Table 1, col. C and F. The reason for the failure is that we do not make use of the actual position of the observed points within the cells. We only ask in which cell they are situated. Although in such a case the tests do not show any difference between the arrangements a_ν and b_λ, some conclusions might be drawn from such 'observations'. If the number of points falling in the first few cells and also in the last few cells is disproportionately large, and if there is a deficiency in the middle cells (Table 1, col. D), we have to conclude that the distribution h_0 is too concentrated or that we have chosen too small a value for the constant which depends only on the standard deviation. If the number of points in the cells at either end is small (Table 1, col. E), the inverse inferences follow. These considerations may give a hint about the choice of an alternative hypothesis.

To illustrate the above methods, let us take the fictitious example given by Pearson (1938) in his Fig. 2, p. 136. He arranges $n = 10$ points in $k = 10$ cells and considers six sets A, B, ..., F, given in Table 1.

Let us suppose that these six sets are the results of six different hypotheses applied to the same observations. The χ^2 test leads to

$$P_A > P_D > P_C = P_F > P_B = P_E.$$

The probabilities of the various columns give the same ordering

$$\Pi_{\text{max.}} = \Pi_A > \Pi_D > \Pi_C = \Pi_F > \Pi_B = \Pi_E.$$

The most probable set contains one point in each cell (set A). It is not possible to decide whether C is more probable than F, and whether B is more probable than E.

Table 1. *Pearson's set*

Class	A	B	C	D	E	F
0·0–0·1	1	2	0	2	0	0
0·1–0·2	1	3	0	1	0	1
0·2–0·3	1	2	0	2	1	2
0·3–0·4	1	1	0	0	2	2
0·4–0·5	1	0	1	0	2	0
0·5–0·6	1	1	2	1	3	1
0·6–0·7	1	0	1	0	1	0
0·7–0·8	1	1	2	1	1	0
0·8–0·9	1	0	2	1	0	2
0·9–1·0	1	0	2	2	0	2
χ^2	0	10	8	6	10	8
P	1	0·350	0·534	0·740	0·350	0·534
Π	1	$\frac{1}{24}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{16}$

The χ^2 and the entropy test are based upon the same data. But the results reached are incomplete, as artificialities are introduced by the classification of the 'observations' into the arbitrary cells. The actual position of the points within the cells is not used. The set A shows that these tests may be misleading in still another way. Each cell in set A contains exactly the expected number. But it would be false to conclude that the hypothesis is true, since the actual positions of the points within the cells might differ from the ideal positions.

Let us suppose we know these positions. It might then happen that the difference between the observed and the ideal positions of the points is smaller for a set K than for a set L, even if the differences between the actual and the theoretical number of points is larger for K than for L.

It is now our task to assign a meaning to the term ideal position and to define a measure of the differences between the 'observed' and the ideal set.

3. The mth point test

The ideal position of n points, distributed with uniform probability over the interval 0, 1, is such that the distances between consecutive pairs of points are equal. But there are a number of ways of distributing n points equidistantly over the interval 0, 1. E. S. Pearson, in the preceding note, suggests that

$$\acute{y}_m = \frac{2m-1}{2n} \tag{23}$$

might be used as the ideal position of the mth point. However, as y is a statistical variable, we should represent it by an average, to choose which we must consider the distribution of the mth point.* Any observation chosen at random has the same probability of falling on any position y within the interval. But for the mth point this probability depends upon y and m. The initial distribution w of the variable $W(x) = y$ is constant. According to (8)

$$w(y) = 1 \quad (0 \leqq y \leqq 1).$$

The probability of obtaining a point equal to or less than y is y. According to (2) the distribution of the mth point is

$$\mathfrak{w}_m(y, n) = \binom{n}{m} m y^{m-1}(1-y)^{n-m}. \tag{24}$$

The distribution (24) is of Karl Pearson's Type I. For $m = 1$ (and $m = n$) the distribution will only decrease (increase). The distribution of the mth point is equal to the distribution of the $(n-m+1)$th point. If we replace y by $1-y$

$$\mathfrak{w}_{n-m+1}(1-y, n) = \mathfrak{w}_m(y, n). \tag{24'}$$

The most probable position $\tilde{y}$ of the mth point is given by

$$\frac{n-m}{1-y} = \frac{m-1}{y},$$

which leads to

$$\tilde{y}_m = \frac{m-1}{n-1}. \tag{25}$$

For given values of m the median position can be obtained from the tables of the Incomplete Gamma Function. To find the arithmetic mean $\bar{y}$ and the control curves, it is necessary to have the moments M_k of (24). They are

$$M_k = \binom{n}{m} m \int_0^1 y^{m+k-1}(1-y)^{n-m}\, dy.$$

According to the well-known properties of the Gamma function

$$M_k = \frac{n!}{(m-1)!\,(n-m)!} \frac{(m+k-1)!\,(n-m)!}{(n+k)!} = \frac{n!}{(n+k)!} \frac{(m+k-1)!}{(m-1)!}.$$

* [The distribution of a ranked individual sampled from a rectangular population, and the moments of this distribution, were obtained by Karl Pearson in the first of two papers (1931, p. 390, and 1932) dealing with ranked variates. Ed.]

Therefore
$$M_{k+1} = M_k \frac{m+k}{n+k+1}. \tag{26}$$

For $k = 1$ the arithmetic mean of the mth point is

$$\bar{y}_m = \frac{m}{n+1}, \tag{27}$$

and for $k = 2$ the second moment is

$$M_2 = \bar{y}_m \frac{m+1}{n+2}.$$

Finally, the variance
$$\sigma_m^2 = \frac{m(n-m+1)}{(n+1)^2(n+2)}. \tag{28}$$

These formulae apply also to the cases $m = 1$ and $m = n$. The standard deviation of the mth value may be written

$$\sigma_m = \sqrt{\frac{\bar{y}_m(1-\bar{y}_m)}{n+2}}. \tag{28'}$$

This formula differs slightly from the general expression (3), and leads to an unexpected result: as we approach the centre from either side, the precision of the mth point decreases. The precision of the mth point will be a minimum for $m = \frac{1}{2}n+1$, if n is even, and for $m = \frac{1}{2}(n+1)$ if n is odd. The values of $\sigma_m\sqrt{(n+2)}$ are given in Table 2.

Table 2. *Standard deviation of the mth point*

$\bar{y}_m$	$\bar{y}_m$	$\sigma_m\sqrt{(n+2)}$
0·05	0·95	0·21794
0·10	0·90	0·30000
0·15	0·85	0·35707
0·20	0·80	0·40000
0·25	0·75	0·43301
0·30	0·70	0·45826
0·35	0·65	0·47697
0·40	0·60	0·49000
0·45	0·55	0·49750
0·50	0·50	0·50000

We must now decide whether to use the mean (27) or the mode (25) as the ideal position of the mth point. The modes of the first and of the last points are 0 and 1 respectively, whereas the corresponding means are

$$\bar{y}_1 = \frac{1}{n+1}, \quad \bar{y}_n = 1 - \frac{1}{n+1}.$$

As the 'observations' $W_0(x_1)$ and $W_0(x_n)$ of the first and the last point differ from 0 and 1, the arithmetic mean is to be preferred.

Formula (27) gives the ideal position and therefore the theoretical numbers of points in each cell which can be compared with the 'observed' numbers. This method leads to an improvement of the tests (18) and (21) where the choice of the cells was still arbitrary and where the actual position of each point was not taken into account.

Besides comparing the uniform distribution with the 'observed' position of the points we can use the corresponding cumulative frequency. The probability scale y is plotted as abscissa and m as ordinate. We count the number of points below m. The mean position $\bar{y}$ of the mth point becomes a straight line differing from the diagonal which represents the modal positions. The figure opposite traces, for $n = 20$, the mean, the modal and Pearson's position of the mth point given by (23).

In the same way we plot the 'observed' points obtained by $h_0, h_1, \ldots$. These probability points $W_0(x_m)$, $W_1(x_m)$ will be scattered about the straight line. Usually the area between the observed cumulative frequency curve, the ordinate and the parallel to the abscissa is kept equal to the corresponding area for the theoretical curve. Since for the present problem no constants have to be determined, we have no way of enforcing this equality. The area J bounded by the diagonal straight line through the points with the co-ordinates $m/(n+1)$, m $(m = 1, 2, \ldots, n)$, the length $1/(n+1)$ to $n/(n+1)$ of the abscissa axis and the two parallels to the ordinate axis, is

$$J = \tfrac{1}{2}(n-1). \tag{29}$$

This might differ from the area $J^{(0)}$ of the $n-1$ 'observed' trapezes

$$y_m,\ m,\ m+1,\ y_{m+1} \quad (m = 1, 2, \ldots, n-1).$$

As y_m are the 'observed' points

$$\begin{aligned} J^{(0)} &= \sum_{m=1}^{n-1} (m+\tfrac{1}{2})(y_{m+1}-y_m) \\ &= \sum_{2}^{n} y_m(m-1) - \sum_{1}^{n-1} my_m + \frac{1}{2}\sum_{2}^{n} y_m - \frac{1}{2}\sum_{1}^{n-1} y_m \\ &= ny_n - \sum_{1}^{n} y_m + \tfrac{1}{2}(y_n - y_1). \end{aligned}$$

If we replace each value y_m by its expectation from (27), we get

$$\begin{aligned} \bar{J}^{(0)} &= \frac{1}{n+1}\{n^2 - \tfrac{1}{2}[n(n+1)] + \tfrac{1}{2}(n-1)\} = \frac{n^2-1}{2(n+1)} \\ &= \tfrac{1}{2}(n-1), \end{aligned} \tag{29'}$$

as it ought to be. The 'observed' area is not equal to the theoretical area, but its expectation is. It might happen that the numerical value of $J^{(0)}$ is very close to J as a result of compensating deviations. Therefore this numerical comparison can be used as a test only in connexion with the graph of the 'observed' and theoretical cumulative histogram of the mth points.

To control the agreement between the 'observed' cumulative histogram and the ideal straight line we use two control curves through the points $\bar{y} \mp \sigma$, m, where $\bar{y}$ is given by (27) and σ by (28′). They are traced in the figure for $n = 20$.

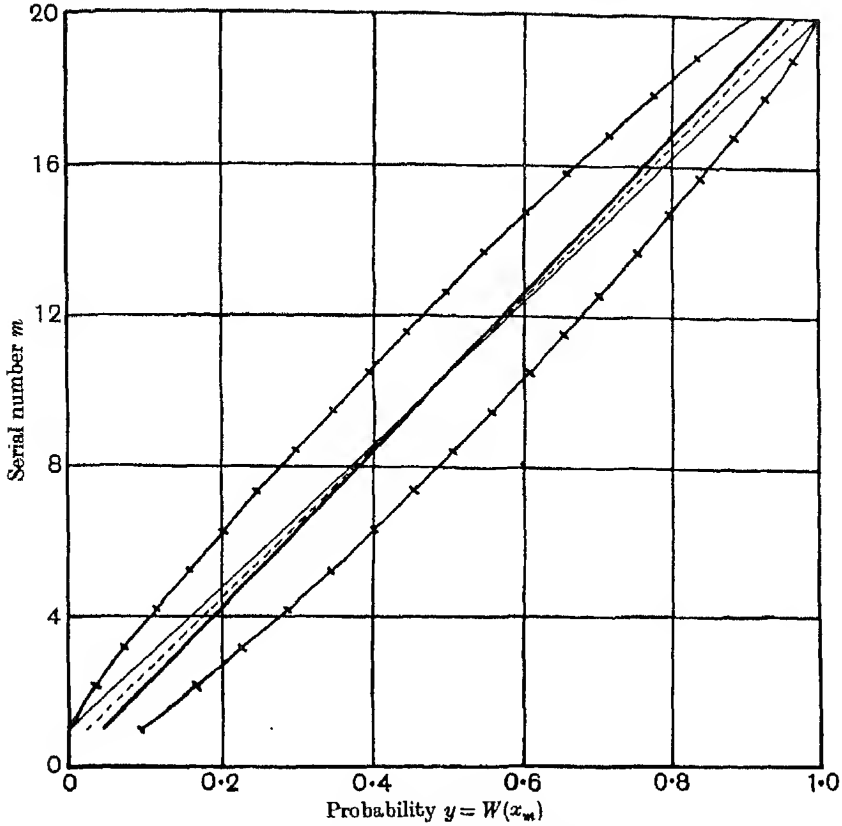


Fig. 1. Control curve for uniform distribution.

——— Mean mth point $\bar{y}$ (27)
– – – – Pearson's mth point $\acute{y}$ (23)
——— Modal mth point $\tilde{y}$ (25)
├─┼─┼─┤ Control curves

It is interesting to calculate the area A bounded by these two curves and to compare it with the area J. If we consider m as abscissa and $\bar{y}+\sigma$ and $\bar{y}-\sigma$ as ordinates we have, for n sufficiently large,

$$A = \int_1^n [\bar{y}+\sigma-(\bar{y}-\sigma)]\,dm$$
$$= 2\int_1^n \sigma\,dm.$$

If we introduce $\bar{y}$ as variable of integration we obtain from (28′)

$$A = \frac{2(n+1)}{\sqrt{(n+2)}} \int_{1/(n+1)}^{n/(n+1)} \sqrt{\{\bar{y}(1-\bar{y})\}}\,d\bar{y}.$$

The transformation

$$\bar{y} = \sin^2 t, \quad 1-\bar{y} = \cos t, \quad d\bar{y} = 2\sin t \cos t\, dt,$$

leads, as is well known, to

$$\frac{\sqrt{(n+2)}}{n+1} A = \tfrac{1}{4}[t - \sin t \cos t]_{2t_0}^{2t_1}.$$

The limits are given by

$$\sin t_0 = \sqrt{\frac{1}{n+1}}, \quad \sin t_1 = \sqrt{\left(1 - \frac{1}{n+1}\right)}.$$

For the expansion of t_0 it is sufficient to put

$$t_0 = \arcsin \sqrt{\frac{1}{n+1}} = \frac{1}{\sqrt{(n+1)}} - \frac{1}{6(n+1)\sqrt{(n+1)}},$$

provided $(n+1)^2 \gg 1$. Under the same condition,

$$\begin{aligned} \arcsin \sqrt{(1-x^2)} &= \arccos x \\ &= \tfrac{1}{2}\pi - \arcsin x, \end{aligned}$$

becomes for any $|x| < 1$

$$\arcsin \sqrt{(1-x^2)} = \tfrac{1}{2}\pi - x + \tfrac{1}{6}x^3.$$

Therefore $$t_1 = \frac{\pi}{2} - \frac{1}{\sqrt{(n+1)}} + \frac{1}{6\sqrt{(n+1)}\,(n+1)}$$

so that $$\frac{t_1 - t_0}{2} = \frac{\pi}{4} - \frac{1}{\sqrt{(n+1)}} + \frac{1}{6(n+1)\sqrt{(n+1)}}.$$

The second factor in the brackets becomes

$$\begin{aligned} 2\sin t_0 \sqrt{(1-\sin^2 t_0)}\,(\cos^2 t_0 - \sin^2 t_0) &= 2\sin t_0 \sqrt{(1-\sin^2 t_0)}\,(1 - 2\sin^2 t_0) \\ &= 2\sqrt{\left\{\frac{1}{n+1}\left(1 - \frac{1}{n+1}\right)\right\}}\left(1 - \frac{2}{n+1}\right) \\ &= \frac{2\sqrt{n}\,(n-1)}{(n+1)^2}. \end{aligned}$$

In the same way

$$\sin 2t_1 \sqrt{(1-\sin^2 2t_1)} = 2\sqrt{\left\{\frac{n}{n+1}\left(1 - \frac{n}{n+1}\right)\right\}}\left(1 - \frac{2n}{n+1}\right) = -\frac{2\sqrt{n}\,(n-1)}{(n+1)^2}.$$

Finally, we reach the area bounded by the control curves

$$A = \frac{n+1}{\sqrt{(n+2)}}\left(\frac{\pi}{4} + \frac{(n-1)\sqrt{n}}{(n+1)^2} - \frac{1}{\sqrt{(n+1)}} + \frac{1}{6(n+1)\sqrt{(n+1)}}\right).$$

According to (29) the ratio of the area bounded by the control curves to the area of the cumulative histogram

$$\frac{A}{J} = \frac{n+1}{n-1}\frac{2}{\sqrt{(n+2)}}\left(\frac{\pi}{4} + \frac{(n-1)\sqrt{n}}{(n+1)^2} - \frac{1}{\sqrt{(n+1)}} + \frac{1}{6(n+1)\sqrt{(n+1)}} \cdots\right), \tag{30}$$

converges towards zero as $1 : \sqrt{n}$.

The properties (29′) and (30) of the cumulative histogram of the positions of the mth points allow for the comparison of the 'observations' $W_0(x_m)$ and the ideal points $m/(n+1)$. It will often be sufficient to inspect the deviations between the 'observations' and theory to judge which set of 'observed' points is closer, on the whole, to the theoretical positions. It seems legitimate to prefer a hypothesis h_0 if the control area contains more points for h_0 than for h_1.

In order to secure a numerical test, we can introduce the mean $\mathfrak{S}^2$ of the sum of the squares of the differences between the 'observed' values $y_m = W_0(x_m)$ and the mean positions $\bar{y}_m = m/(n+1)$. Take

$$\mathfrak{S}^2 = \frac{k}{n}\sum_{m=1}^{n}\left(y_m - \frac{m}{n+1}\right)^2, \tag{31}$$

where the value of the constant k will be specified later. One extreme for (31) would be to have the theory hold for every point. Then the value of the sum would be zero. The other extremes would be when all points are concentrated either at the origin, zero, or at the end, 1. In the first case

$$\frac{1}{n(n+1)^2}\sum_{m=1}^{n} m^2 = \frac{2n+1}{6(n+1)}.$$

The second case leads to the same value, since $\Sigma m = \frac{1}{2}n(n+1)$ and therefore

$$\frac{1}{n}\sum_{m=1}^{n}\left(1 - \frac{m}{n+1}\right)^2 = 1 - 1 + \frac{1}{n(n+1)^2}\sum_{m=1}^{n} m^2.$$

Therefore
$$0 \leqslant \mathfrak{S}^2 \leqslant k\frac{2n+1}{6(n+1)} < \frac{k}{3}. \tag{32}$$

In order to draw conclusions from an observed value $\mathfrak{S}^2$ we have to calculate its expectation $\overline{\mathfrak{S}^2}$. We will determine k in such a way that $\overline{\mathfrak{S}^2}$ is independent of n.

A test similar to (31), but serving another purpose, has been introduced for the usual distributions by H. Cramér (1928) and R. von Mises (1931). When applied to uniform distributions, this ω^2 test leads to the use of $\acute{y}_m$ of (23) instead of the mean value $\bar{y}_m$. For this test the sum of the deviations is zero which does not hold in our case. The expectation $\overline{\mathfrak{S}^2}$ of $\mathfrak{S}^2$ is

$$\overline{\mathfrak{S}^2} = \frac{k}{n}\left(\sum_{m=1}^{n}\left[\overline{y_m^2} - \frac{2m\bar{y}_m}{n+1} + \frac{m^2}{(n+1)^2}\right]\right).$$

The first two sums are obtained from (27) and (26). Therefore

$$\overline{\mathfrak{S}^2} = \frac{k}{n}\left(\frac{\sum_1^n m(m+1)}{(n+1)(n+2)} - \frac{\sum_1^n m^2}{(n+1)^2}\right)$$

$$= \frac{k}{n(n+1)}\left(\frac{\sum_1^n m}{n+2} - \frac{\sum_1^n m^2}{(n+1)(n+2)}\right).$$

The introduction of the sum of the powers of the natural numbers leads to

$$\overline{\mathfrak{S}^2} = \frac{k}{2(n+2)}\left(1-\frac{2n+1}{3n+3}\right) = \frac{k}{6(n+1)}.$$

Taking $k = 6(n+1)$ we propose therefore, as test of a hypothesis h_0, the coefficient

$$\mathfrak{S}^2 = \frac{6(n+1)}{n}\sum_{m=1}^{n}\left(y_m - \frac{m}{n+1}\right)^2, \tag{33}$$

which, according to (32), can assume values between zero and $2n+1$, and has for expectation the value

$$\overline{\mathfrak{S}^2} = 1. \tag{34}$$

Of two competing hypotheses, the one with the smaller value of $\mathfrak{S}^2$ is to be preferred.

The $\mathfrak{S}^2$ criterion does not introduce any arbitrary classification. It makes use of all observations. Besides the probabilities $W(x_m)$ corresponding to the observed values x_m no new calculations are needed. The test has a clear meaning and its application is simple. This is due to the fact that it is a natural consequence of the probability integral transformation.

Summary

We propose the following procedure for testing statistical hypotheses: The constants for competing hypotheses, having the same number of constants, are determined in such a way that their precisions are approximately the same. Then we calculate the probabilities $W_0(x_m)$, $W_1(x_m)$, ..., and their respective control curves. We trace $W(x_m)$, $x_m \mp \sigma_m$ and compare it with the observed frequency curve. If neither the classical tests nor the control curves indicate a clear superiority of one of the hypotheses we consider the probabilities as 'observations' and plot the corresponding points on the y axis. We now compare, by formulae (9)—(17), the 'observations' with the theoretical uniform distribution and apply the χ^2 and entropy test of formulae (18) and (21), respectively. If necessary, we repeat these tests in such a way that the actual position of each point is taken into account. Formula (18′) gives a value of χ^2 which is independent of the classification. Then we plot the cumulative frequency of the 'observations', which is compared with the straight line (27) and controlled by the values given in Table 2. The final test is given by (33).

REFERENCES

Cramér, H. (1928). On the composition of elementary errors. *Skand. Aktuar. Tidskr.* **11**, 13–17, 141–80.

Gumbel, E. J. (1935). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré*, **4**, fasc. 2, 115–58.

—— (1936). La précision de la moyenne arithmétique et de la médiane. *Aktuar. Vedy*, **6**, no. 4, 145–54.

—— (1937). Les intervalles extrêmes entre les émissions radioactives. I. *J. Phys. Radium*, série vii, **8**, no. 8, pp. 321–29; ii, 446–52.

—— (1939). Les valeurs de position d'une variable aléatoire. *C.R. Acad. Sci., Paris*, **208**, 147–9.

Mises, R. von (1931). *Wahrscheinlichkeitsrechnung*, 316–35. Leipzig.

Neyman, J. (1937). 'Smooth test' for goodness of fit. *Skand. Aktuar. Tidskr.* 149–99.

Neyman, J. & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20** A, 263–94.

Pearson, E. S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biometrika*, **30**, 134–48.

Pearson, Karl (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, **25**, 379–410.

—— (1934). On a new method of determining goodness of fit. *Biometrika*, **26**, 425–42.

Pearson, Karl & Pearson, M. V. (1931, 1932). On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. Part I: *Biometrika*, **23**, 364–97. Part II: *Biometrika*, **24**, 203–79.

THE RANGE IN RANDOM SAMPLES

By H. O. HARTLEY

1. Introduction

If the observations x_ν $(\nu = 1, 2, \ldots, n)$ of a random sample are arranged in ascending order of magnitude $(x_{\nu+1} > x_\nu)$ the range w in such samples is defined as the distance between the two extreme observations

$$w = x_n - x_1.$$

It may therefore be regarded as a measure of the variability or dispersion among the observations of the sample. Theoretically its efficiency in the sense defined by R. A. Fisher is, as a rule, much inferior to that of the standard deviation. Moreover, extensive investigations have shown that its random sampling distribution is markedly dependent on the parental population (E. S. Pearson, 1926). For large samples x_ν drawn from a parental distribution $f(x)$ the extreme values x_1 and x_n will lie right inside the lower and upper tail of $f(x)$, and in practice it is only in exceptional cases that the exact shape of $f(x)$ has been established to such a degree of accuracy that the resulting distribution of w can be trusted for large n. In most cases the use of the range must therefore be limited to small samples, say with $2 \leqslant n \leqslant 20$.

Large numbers of small samples may often be used with advantage when the mean range is calculated as an estimate of the standard deviation of the population (Pearson & Haines, 1935). Although theoretically such an estimate is not efficient and certainly not sufficient, it is nevertheless of considerable importance in many fields of application because of its simplicity. Statistical control charts in industrial quality control make extensive use of it, and more recently the range has been applied to investigations in gunnery.

In some fields of application a disadvantage may arise from the fact that the range is an inexact statistic; its random sampling distribution depends on the standard deviation of the parent. This applies in particular to the analysis of small samples in biological experiments. The tendency of modern small sample theory has been to replace such statistics by what are called exact statistics, obtained by substituting for the unknown standard deviation of the parent an estimate calculated from an independent sample. This particular process of reaching exact statistics has sometimes been referred to as 'Studentization'. A general theory of this process will be given in a further paper which it is hoped to publish in this journal, where it will be shown how estimates of scale parameters in general, and of the range in particular, may be converted into exact statistics. In this paper, however, we deal with the case where the standard deviation of the parent is known. Indeed, it is this dependence of the random sampling distribution of range on the scale parameter of the parent which makes it possible to estimate its efficiency as an estimate theoreof.

The question of grouping has been a subject of investigation in the case of the sample standard deviation; we shall here deal with the effect of grouping on the range, a problem which has so far received, we believe, no attention whatsoever.* As practical examples of the occurrence of grouping we may quote three instances:

(*a*) The rounding off of data for convenience of recording and analysis.

(*b*) The recording of data to the nearest unit of measurement. Where the technique is of low accuracy (see e.g. Tildesley, 1940) the unit of measurement will be comparable in magnitude with the standard deviation of the actual data.

(*c*) The analysis of data which are classified in categories. In such cases we may often find that the original data are unobtainable so that group frequencies are the only material available for an analysis.

It will also be shown how the random sampling distribution of the range in grouped samples provides a suitable approach to that of the true range (ungrouped range in sample) on which extensive work has been done in earlier papers published in *Biometrika*. The mathematical formulae developed in this paper make this complex distribution amenable to a tabulation. For the case of normal samples, the work has actually been carried out and the resulting tables of the probability integral are given and discussed elsewhere in the present issue of this journal.

2. THE DISTRIBUTION OF THE RANGE IN A GROUPED SAMPLE

Let us denote by $x_1, \ldots, x_n$ the observations in a random sample drawn from the parental distribution $f(x)$† and arranged in ascending order of magnitude. This sample is now classified in groups or categories of constant length h with equidistant end-points

$$\ldots, \xi-kh, \ldots, \xi-h, \xi, \xi+h, \ldots, \xi+kh, \ldots, \tag{1}$$

covering the whole x scale from $-\infty$ to $+\infty$. Let us denote by ξ_n and ξ_1 the respective *centres* of the categories containing x_n and x_1. Then the problem is to find the random sampling distribution of the range in a grouped sample, i.e. of $\xi_n-\xi_1$. The mean of this distribution is of particular interest. Obviously this statistic can only assume values which are multiples of the group interval h and is therefore discontinuous. Like the distribution of the 'ungrouped' or true range it depends on the standard deviation σ of the parent $f(x)$. In addition, it depends

* The effect of grouping seems to be of some importance in researches on the technique of anthropological measurement (Tildesley, 1940), where some of the results given below have already been applied before this paper had gone to press.

† We shall deal here with a parental population represented by a 'piecewise continuous' distribution function $f(x)$. A function is called 'piecewise continuous' for $-\infty<x<+\infty$ if in any closed interval of x the function $f(x)$ is continuous apart from a finite number of ordinary discontinuities. If the actual range of the variate is bounded we simply define $f(x)\equiv 0$ outside this range. Moreover, we assume that $f(x)$ has contact of at least second order at $\pm\infty$. It is easy to see how our results may be generalized to cover distribution functions with singularities.

on the category width h and on the position of the category midpoints relative to the population mean X. Of these parameters only h will in practice be known. Methods to eliminate σ are to be given in a separate paper whilst the elimination of X is dealt with in the section on randomized grouping (6).

It will be convenient to use the following notation:

$$\int_i^j = \int_{\xi+ih}^{\xi+jh} f(x)\,dx, \quad \int_{-\infty}^i = \int_{-\infty}^{\xi+ih} f(x)\,dx, \quad \int_i^\infty = \int_{\xi+ih}^{\infty} f(x)\,dx.$$

Let us now find the chance that $\quad \xi_n - \xi_1$

is at most $(m-1)h$, and that in addition

$$\xi_1 = \xi + (i+\tfrac{1}{2})h$$

for a particular value of i. This chance is given by

$$\left(\int_i^{i+m}\right)^n - \left(\int_{i+1}^{i+m}\right)^n. \tag{2}$$

The first term in (2) represents the probability for all x_ν to lie between $\xi+ih$ and $\xi+(i+m)h$. From this we have to deduct the chance for all x_ν to lie between $\xi+(i+1)h$ and $\xi+(i+m)h$ which is given by the second term of (2). In taking the difference we are therefore left with the chance for the occurrence of a sample completely contained in the interval $\xi+ih$ to $\xi+(i+m)h$ but with at least one of the x_i lying between $\xi+ih$ and $\xi+(i+1)h$. This proves that (2) represents the required chance. Now, since all samples may be classified with regard to their lowest category, the probability for $\xi_n - \xi_1$ to be at most $(m-1)h$ is given by summation over all i of the expression (2). If we denote this probability by $P(n, h, m-1, \xi)$ we find

$$P(n, h, m-1, \xi) = \sum_{i=-\infty}^{+\infty} \left\{\left(\int_i^{i+m}\right)^n - \left(\int_{i+1}^{i+m}\right)^n\right\}. \tag{3}$$

With equation (3) we have reached a formal representation of the random sampling distribution of $\xi_n - \xi_1$. Its evaluation is a simple matter for large group intervals h and for parental distributions $f(x)$ with a tabulated probability integral $\int f(x)\,dx$. If we were to take the trouble of tabulating the probability integral (3) we should obtain the mean of $\xi_n - \xi_1$ as a by-product from a summation of (3). It will be shown in the next section that this summation, if carried out analytically, produces a very simple formula for this mean.

3. The mean range in a grouped sample

To find the mean of the distribution it is convenient to extend the summation in (3) from some finite negative value $i = -j$ up to $+\infty$. By choosing j sufficiently large the resulting error may be made negligible. We introduce

$$p_{-j}(m-1) = \sum_{i=-j}^{\infty} \left[\left(\int_i^{i+m}\right)^n - \left(\int_{i+1}^{i+m}\right)^n\right], \tag{4}$$

and find for the difference between $P(m-1)$ and $p_{-j}(m-1)$*

$$|P(m-1)-p_{-j}(m-1)| \leqslant \sum_{i=-\infty}^{-j} n\int_{i}^{i+1} = n\int_{-\infty}^{-j+1}, \tag{5}$$

for all m and j. To find the mean of $\xi_n - \xi_1$ we must first note the probability for this statistic to be exactly equal to mh, where $m = 0, 1, 2, \ldots$. Denoting this probability by $\phi(m)$ we have from the definition of $P(m)$

$$\phi(m) = P(m) - P(m-1).$$

If we denote the mean of $\xi_n - \xi_1$ by Ξ we have by definition

$$\Xi = h\sum_{k=0}^{\infty} \phi(k)\,k$$

$$= h \lim_{m\to\infty} \{(m+1)\,P(m) - S_m\}, \tag{6}$$

where $$S_m = (m+1)\,\phi(0) + m\phi(1) + \ldots + \phi(m)$$

or $$S_m = P(0) + P(1) + \ldots + P(m). \tag{7}$$

To find Ξ let us first consider the second term in formula (6). We have from equations (7) and (5)

$$S_m = p_{-j}(0) + \ldots + p_{-j}(m) + \epsilon_1, \tag{8}$$

where $$|\epsilon_1| \leqslant n(m+1)\int_{-\infty}^{-j+1}, \tag{9}$$

for all j and m. Now from the definition of $p_{-j}(k)$ we find

$$\sum_{k=0}^{m} p_{-j}(k) = \sum_{k=0}^{m}\left\{\sum_{i=-j}^{\infty}\left[\left(\int_{i}^{i+k+1}\right)^n - \left(\int_{i+1}^{i+k+1}\right)^n\right]\right\}$$

$$= \sum_{k=0}^{m}\left\{\left(\int_{-j}^{-j+k+1}\right)^n + \sum_{i=-j}^{\infty}\left[\left(\int_{i+1}^{i+k+2}\right)^n - \left(\int_{i+1}^{i+k+1}\right)^n\right]\right\}$$

$$= \sum_{k=0}^{m}\left(\int_{-j}^{-j+k+1}\right)^n + \sum_{i=-j}^{\infty}\left(\int_{i+1}^{i+m+2}\right)^n.$$

Putting now $m = 2j$, we have

$$\sum_{k=0}^{2j} p_{-j}(k) = \sum_{i=-j+1}^{j+1}\left(\int_{-\infty}^{i}\right)^n + \sum_{i=-j+1}^{j+1}\left(\int_{i}^{\infty}\right)^n + \epsilon_2 + \epsilon_3, \tag{10}$$

where it is easily seen that

$$|\epsilon_2| \leqslant 2jn\int_{-\infty}^{-j}, \quad |\epsilon_3| \leqslant 2jn\int_{j+2}^{\infty} + \sum_{i=j+1}^{\infty}(i-j)\int_{i}^{i+1}. \tag{11}$$

Finally, we want to replace in formula (6) the first term

$$(2j+1)\,P(2j) \quad \text{by} \quad (2j+1). \tag{12}$$

* In this section we deal with fixed group intervals and a fixed sample size n so that we drop the arguments n, h and ξ.

The resulting error is easily estimated. We have from (5)

$$(2j+1)(1-P(2j)) = (2j+1)(1-p_{-j}(2j))+\epsilon_4$$

where
$$|\epsilon_4| \leqslant (2j+1)\,n\int_{-\infty}^{-j+1}. \tag{13}$$

Moreover, according to the definition (4) we may write

$$\begin{aligned}(2j+1)\,p_{-j}(2j) &= (2j+1)\left\{\sum_{i=-j}^{\infty}\left(\left(\int_{i}^{i+2j+1}\right)^n - \left(\int_{i+1}^{i+2j+2}\right)^n\right)\right\} \\ &\quad + (2j+1)\left\{\sum_{i=-j}^{\infty}\left(\left(\int_{i+1}^{i+2j+2}\right)^n - \left(\int_{i+1}^{i+2j+1}\right)^n\right)\right\} \\ &= (2j+1)\left(\int_{-j}^{+j+1}\right)^n + \epsilon_5,\end{aligned}$$

where
$$|\epsilon_5| \leqslant (2j+1)\sum_{i=-j}^{\infty} n\int_{i+2j+1}^{i+2j+2} \leqslant n(2j+1)\int_{j+1}^{\infty}, \tag{14}$$

so that finally we have

$$(2j+1)(1-p_{-j}(2j)) = -\epsilon_5+\epsilon_6,$$

with
$$\epsilon_6 \leqslant (2j+1)\,n\left(\int_{j+1}^{\infty}+\int_{-\infty}^{-j}\right). \tag{15}$$

The error terms ϵ_1, ϵ_2, ϵ_3, ϵ_4, ϵ_5 and ϵ_6 are of the form

$$cj\int_{j}^{\infty}; \quad cj\int_{-\infty}^{-j} \quad \text{or} \quad \sum_{i=j}^{\infty}(i-c)\int_{i}^{i+1}.$$

It is easy to see that the above terms tend to 0 as $j \to \infty$. To prove this for the first term we write

$$j\int_{j}^{\infty} \leqslant \left(\frac{\xi+jh}{h}+c\right)\int_{j}^{\infty} < c\int_{j}^{\infty}+\frac{1}{h}\int_{\xi+jh}^{\infty} xf(x)\,dx,$$

which tends to 0 as $f(x)$ has contact of order 1 at $+\infty$. The proof for the other terms is identical. For sufficiently large j we can therefore use the approximations given by (10) and (12) and transform equation (6) into the convenient form

$$\Xi = h\lim_{j\to\infty}\left\{(2j+1) - \sum_{i=-j+1}^{j+1}\left(\int_{-\infty}^{i}\right)^n - \sum_{i=-j+1}^{j+1}\left(\int_{i}^{\infty}\right)^n\right\}. \tag{16}$$

Equation (16) gives the mean range Ξ in a grouped sample in terms of powers of the probability integral of the parental population. For a normal distribution this is a particularly simple formula since such powers have already been calculated by L. H. C. Tippett (1925) and are conveniently tabulated in Table XXI of the *Tables for Statisticians and Biometricians*, Part II. A table of Ξ can, therefore, be easily computed by adding a few entries from Tippett's table and deducting the sum from the appropriate value of $2j+1$.

This has been done for samples of five, ten and twenty observations grouped in categories of breadth h (see table on p. 339). The parameter ξ denotes the

distance of the population mean $X = 0$ from the nearest group end-point. For given h the mean range Ξ in grouped samples is obviously a symmetrical periodic function of ξ with period h. The table has been extended to cover rather coarse grouping intervals ($h = 2{\cdot}2\sigma$) in order to illustrate the possible bias of range when estimated from frequency tables with as few as two or three categories. It is apparent that for small or moderate group intervals, say $h \leqslant \sigma$, the mean range is practically independent of h and ξ, so that no correction (corresponding to the well-known Sheppard's correction for the sample standard deviation) is required for the

Table of mean range in grouped samples drawn from a normal population having unit standard deviation

Size of sample $= n$. Width of group interval $= h$.
Distance of population mean to nearest group-end point $= \xi$.

h	ξ	$n=5$	$n=10$	$n=20$
0·2	0·0	2·32 593	3·07 751	3·73 495
0·6	0·0	2·32 593	3·07 750	3·73 500
	0·2	2·32 593	3·07 751	3·73 492
1·0	0·0	2·32 532	3·08 122	3·72 917
	0·2	2·32 574	3·07 865	3·73 317
	0·4	2·32 642	3·07 450	3·73 962
1·4	0·0	2·31 042	3·06 204	3·82 069
	0·2	2·31 626	3·06 787	3·78 826
	0·4	2·32 938	3·08 095	3·71 575
	0·6	2·33 990	3·09 143	3·65 796
1·8	0·0	2·29 227	2·90 539	3·67 974
	0·2	2·30 022	2·94 491	3·69 553
	0·4	2·32 023	3·04 621	3·73 085
	0·6	2·34 276	3·16 356	3·76 263
	0·8	2·35 734	3·24 140	3·77 838
2·2	0·0	2·36 011	2·77 080	3·27 509
	0·2	2·35 552	2·81 669	3·34 611
	0·4	2·34 221	2·94 307	3·53 896
	0·6	2·32 265	3·11 581	3·79 662
	0·8	2·30 255	3·28 173	4·03 847
	1·0	2·28 963	3·38 358	4·18 452

range. For $h = 0{\cdot}2\sigma$ the mean range in the grouped sample agrees with the theoretical ungrouped range to five places of decimals (see Table of Mean Range, Table XXII of *Tables for Statisticians and Biometricians*, Part II). For coarse grouping the correction becomes important but depends on ξ (as well as on h). For fixed h, as ξ varies from $-\frac{1}{2}h$ to $+\frac{1}{2}h$ the grouped mean range oscillates about the true mean range as a smooth single-period function. The reason for this is obvious. If ξ has a position such that the average positions of x_n and x_1 both happen to fall within the outside halves of two group intervals, then $\xi_n - \xi_1$ will

on the average, be smaller than $x_n - x_1$; vice versa, if the average positions of x_n and x_1 are in the inside halves of two group intervals, there will be a predominance of samples for which $\xi_n - \xi_1$ is larger than $x_n - x_1$.

Moreover, as h increases, the grouped mean range becomes a less reliable estimate, and it can be shown that for $h > \sigma$ the standard deviation of the random sampling distribution of the grouped range rapidly increases with h.

We are thus led to consider two problems; one is the elimination of the parameter ξ (or the dependence of the distribution on the position of the parental population mean); the other is to investigate more closely the random sampling distribution of the grouped range. Before dealing with these problems, however, we must first consider the distribution of the true range (range in the ungrouped sample).

4. The probability integral of the range in random samples

As before we denote by $x_1, \ldots, x_n$ the observations in a random sample drawn from a parental distribution $f(x)$ and arranged in ascending order of magnitude. The range in such a sample, defined as $w = x_n - x_1$, may be regarded as the limit of the grouped range $\xi_n - \xi_1$ as h, the group interval, tends to 0, i.e.

$$w = \lim_{h \to 0} \xi_n - \xi_1.$$

The probability integral of the range w, denoted by $P_n(W)$, is therefore the limit of $P(n, h, m-1, \xi)$, given by (3), as h tends to 0. To obtain this limit we write equation (3) as follows:

$$\begin{aligned} P(n, h, m-1, \xi) &= \sum_{i=-\infty}^{+\infty} \left(\int_{\xi+ih}^{\xi+(i+m)h} f(x)\,dx \right)^n - \left(\int_{\xi+(i+1)h}^{\xi+(i+m)h} f(x)\,dx \right)^n \\ &= \sum_{i=-\infty}^{+\infty} n \left(\int_{\bar{\xi}_i}^{\xi+(i+m)h} f(x)\,dx \right)^{n-1} f(\bar{\xi}_i)\,h, \end{aligned}$$

where $\bar{\xi}_i$ is some mean value between $\xi + ih$ and $\xi + (i+1)h$.

We now put $\qquad m = W/h \quad \text{or} \quad W = mh,$

and let h tend to 0, m to ∞, keeping W constant. We obtain without difficulty

$$\begin{aligned} P_n(W) &= \lim_{\substack{h \to 0 \\ m \to \infty}} P(n, h, m-1, \xi) \\ &= n \int_{-\infty}^{+\infty} f(\xi) \left(\int_{\xi}^{\xi+W} f(x)\,dx \right)^{n-1} d\xi, \end{aligned} \tag{17}$$

which is the required probability integral of the range. This integral may be compared with the expression for the distribution function of w which was given by A. T. McKay & E. S. Pearson (1933). It is easily verified that the function $\phi(w)$ given by these authors is the differential of $P(W)$. The expression for $P(W)$

is decidedly simpler than that for $\phi(w)$ which was used by Pearson for numerical work on this function. However, even $P(W)$ is of a complex character, and only in special cases is it possible to evaluate it analytically. For the rectangular distribution function ($f(x) \equiv 1$, $0 \leqslant x \leqslant 1$) this can be done easily.

5. THE PROBABILITY INTEGRAL OF THE RANGE IN SAMPLES FROM A NORMAL POPULATION

Of particular interest is the case where the parental population is normal. In this case we have

$$f(x) = z(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \tag{18}$$

L. H. C. Tippett (1925), E. S. Pearson (1926, 1932) and A. T. McKay & E. S. Pearson (1933) have considered this problem and carried out extensive numerical work. The method adopted was to calculate correct values of the means and standard deviations of the distributions (as functions of n) and then, with the help of approximate values of β_1 and β_2 use as approximations to the unknown true distribution Pearson-type curves fitted by the method of moments. The numerical results, although they have been successfully tested by experimental sampling, have, of course, an unknown accuracy. It is therefore desirable to find a method which produces $P_n(W)$ to known and sufficient accuracy.

We have
$$P_n(W) = n\int_{-\infty}^{+\infty} z(\xi)\left(\int_{\xi}^{\xi+W} z(x)\,dx\right)^{n-1} d\xi$$
$$= n\int_{-\infty}^{-\frac{1}{2}W} + n\int_{-\frac{1}{2}W}^{+\infty} = I_1 + I_2 \quad \text{(say)}.$$

Writing
$$\eta = -(\xi + W), \quad -\xi = \eta + W,$$

we obtain
$$P_n(W) = n\int_{-\frac{1}{2}W}^{\infty} z(-\eta - W)\left(\int_{-(\eta+W)}^{-\eta} z(x)\,dx\right)^{n-1} d\eta + I_2.$$

Using the symmetry of $z(x)$ we may write

$$P_n(W) = n\int_{-\frac{1}{2}W}^{\infty} z(\eta + W)\left(\int_{\eta}^{\eta+W} z(x)\,dx\right)^{n-1} d\eta + I_2,$$

and writing ξ as a variable of integration in place of η we find

$$P_n(W) = n\int_{-\frac{1}{2}W}^{\infty} [z(\xi + W) + z(\xi)]\left(\int_{\xi}^{\xi+W} z(x)\,dx\right)^{n-1} d\xi,$$

or
$$P_n(W) = -n\int_{-\frac{1}{2}W}^{\infty} [z(\xi + W) - z(\xi)]\left(\int_{\xi}^{\xi+W} z(x)\,dx\right)^{n-1} d\xi$$
$$+ 2n\int_{-\frac{1}{2}W}^{\infty} z(\xi + W)\left(\int_{\xi}^{\xi+W} z(x)\,dx\right)^{n-1} d\xi. \tag{19}$$

Integrating the first integral and introducing $u = \xi + W$ in the second integral we finally obtain

$$P_n(W) = \left(\int_{-\frac{1}{2}W}^{+\frac{1}{2}W} z(x)\,dx\right)^n + 2n\int_{\frac{1}{2}W}^{\infty} z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^{n-1} du. \tag{20}$$

For large values of W this is an approximate solution of the problem since the second term in (20) is small, so that the first term

$$\left(\int_{-\frac{1}{2}W}^{+\frac{1}{2}W} z(x)\,dx\right)^n$$

gives a fair approximation to $P_n(W)$. This expression denotes the chance of observing samples with observations all lying between $-\frac{1}{2}W$ and $+\frac{1}{2}W$; all these samples have a range smaller than or equal to W. For large W it is these samples which constitute an ever-increasing proportion of the total number of samples with range $\leqslant W$.

The second term in (20), which is always positive, takes into account all those samples which are not contained in the interval $-\frac{1}{2}W$ to $+\frac{1}{2}W$. This term cannot be ignored if high accuracy is required and if W is small or moderate. Nevertheless, the work involved in the numerical integration has been considerably reduced, for the range of integration is now from $+\frac{1}{2}W$ to $+\infty$.

The numerical integration of

$$\int_{\frac{1}{2}W}^{\infty} z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^{n-1} du$$

is best carried out simultaneously for values of n forming an arithmetical progression. For fixed u and W, the integrand is then given by the terms of a geometrical progression with, say,

$$z(u)\left(\int_{u-W}^{u} z(x)\,dx\right)^2$$

as first term and

$$\left(\int_{u-W}^{u} z(x)\,dx\right)$$

as common ratio. Such a geometrical progression can be produced automatically by certain modern calculating machines. This forms the basic idea of the actual computation of the probability integral which is described in detail in another paper (pp. 309–10 above).

6. The range in randomly grouped samples

The results of sections (2) and (3) on the effect of grouping on the distribution of range depend on the parameter ξ, which denotes the origin of the equidistant set of group end-points

$$\xi + ih \quad i = 0, 1, 2, \ldots,\; -1, -2, \ldots,$$

given by equation (1). In practice, however, all we know is the category breadth, h.

We then select the actual group end-points $\xi + ih$ from considerations which are, by necessity, independent of the position of the mean X of the parental population $f(x)$, because this position will generally be unknown. One of our group end-points, however, is bound to fall into the interval

$$X - \tfrac{1}{2}h \quad \text{to} \quad X + \tfrac{1}{2}h,$$

wherever this interval may lie. Now, since the origin ξ in our system of group end-points is wholly arbitrary we may assume that for given h we have the inequality

$$X - \tfrac{1}{2}h \leqslant \xi \leqslant X + \tfrac{1}{2}h.$$

The fact that our group end-points (and therefore their origin ξ) are chosen independently of X has now to be expressed in mathematical terms. This is done by assuming that we are dealing with a population of values of ξ (being the origins of corresponding systems of group end-points) which are rectangularly distributed in the interval $X - \tfrac{1}{2}h \leqslant \xi \leqslant X + \tfrac{1}{2}h$. This condition is exactly fulfilled where grouping has been introduced through rounding off of data (example (a) on p. 335), and it is often an appropriate assumption in the other examples, as in many other cases of grouping which occur in practice.*

In order to derive the distribution of range in samples randomly grouped in the above sense we have to return to section (2). In this section we derived the probability $P(n, h, m-1, \xi)$, giving the chance that the difference between the centre points ξ_n and ξ_1 of the highest and lowest category covered by a sample of n items is at most $(m-1)h$, where h is the constant category breadth and group end-points are given by (1). The frequency distribution of $\xi_n - \xi_1$ which we may denote by $\phi(n, h, m, \xi)$ is then given by

$$\phi(n, h, m, \xi) = P(n, h, m, \xi) - P(n, h, m-1, \xi),$$

and represents the chance that $\xi_n - \xi_1$ is exactly mh, given a particular value of ξ. The corresponding frequency distribution for random grouping may be denoted by $\phi(n, h, m)$. To derive it we may apply Bayes's Theorem and obtain

$$\phi(n, h, m) = \frac{1}{h}\int_{X-\frac{1}{2}h}^{X+\frac{1}{2}h} \phi(n, h, m, \xi)\, d\xi.$$

The resulting cumulative probability may therefore be defined by

$$P(n, h, m) = \sum_{j=1}^{m} \phi(n, h, j) = \frac{1}{h}\int_{X-\frac{1}{2}h}^{X+\frac{1}{2}h} \sum_{j=1}^{m} \phi(n, h, j, \xi)\, d\xi,$$

which yields the corresponding relation for the cumulative probabilities

$$P(n, h, m) = \frac{1}{h}\int_{X-\frac{1}{2}h}^{X+\frac{1}{2}h} P(n, h, m, \xi)\, d\xi.$$

* In certain cases, when grouping is very coarse, it may be advantageous to use an estimate $\bar{x}$ of the population mean X (either dependent or independent of the sample whose range is considered). The increase in information is akin to that given by an ancillary statistic in estimation theory. However, from the results in § 3 it would appear that little information is gained where the grouping interval is small or moderate.

Substituting, now, the expression (3) for $P(n,h,m,\xi)$ we obtain

$$P(n,h,m)=\frac{1}{h}\sum_{i=-\infty}^{+\infty}\int_{X-\frac{1}{2}h}^{X+\frac{1}{2}h}\left\{\left(\int_{\xi+ih}^{\xi+(i+m+1)h}f(x)\,dx\right)^n-\left(\int_{\xi+(i+1)h}^{\xi+(i+m+1)h}f(x)\,dx\right)^n\right\}d\xi$$

$$=\frac{1}{h}\int_{-\infty}^{+\infty}\left\{\left(\int_{\xi}^{\xi+(m+1)h}f(x)\,dx\right)^n-\left(\int_{\xi}^{\xi+mh}f(x)\,dx\right)^n\right\}d\xi. \tag{21}$$

This formula may be reduced to a simpler form in which its relation to the probability integral of the true range (17) becomes apparent. We introduce the second integral

$$F_n(W)=\int_0^W P_n(w)\,dw$$

$$=\int_0^W n\int_{-\infty}^{+\infty}f(\xi)\left(\int_{\xi}^{\xi+w}f(x)\,dx\right)^{n-1}d\xi\,dw. \tag{22}$$

The first integration is with regard to w and the integrand is an integral with regard to ξ. If, now, in this latter integral $\eta=\xi+w$ is used as variable of integration in place of ξ we have

$$F_n(W)=\int_0^W n\int_{-\infty}^{+\infty}f(\eta-w)\left(\int_{\eta-w}^{\eta}f(x)\,dx\right)^{n-1}d\eta\,dw.$$

We now note that the integrand may be written as a differential with regard to w. Thus, interchanging the order of integration we obtain

$$F_n(W)=\int_{-\infty}^{+\infty}\int_0^W\frac{d}{dw}\left\{\int_{\eta-w}^{\eta}f(x)\,dx\right\}^n dw\,d\eta$$

$$=\int_{-\infty}^{+\infty}\left\{\int_{\eta-W}^{\eta}f(x)\,dx\right\}^n d\eta, \tag{23}$$

thus eliminating integration with regard to w. The surviving variable of integration η may now be replaced by $\xi=\eta-W$ so that we reach the final result

$$F_n(W)=\int_{-\infty}^{+\infty}\left(\int_{\xi}^{\xi+W}f(x)\,dx\right)^n d\xi. \tag{24}$$

We now observe that this integral is identical with the one occurring in the expression (21) for $P(n,h,m)$, and we note the relation

$$P(n,h,m)=\frac{1}{h}\{F_n(\overline{m+1}\,h)-F_n(mh)\}$$

$$=\frac{1}{h}\int_{mh}^{(m+1)h}P_n(w)\,dw. \tag{25}$$

This simple formula makes it possible to obtain the effect of random grouping on the probability integral of range $P_n(W)$. In particular, for normal samples for which $P_n(W)$ has been tabulated at the fine interval of $\Delta w=0{\cdot}05$, the second integral $F_n(W)$ is easily obtained numerically by summation of the tabular entries in the table on pp. 302–7 above.

For $h \to 0$, $m \to \infty$, $mh \to W$, we note that, as expected, $P(n,h,m) \rightrightarrows P_n(W)$; that is, as grouping becomes finer and finer the probability integral of the grouped range tends to that of the true range. What is not quite obvious, however, is the identity of mean grouped range and mean true range, no matter how large the breadth, h, of the randomly placed groups. This point we shall now examine.

If we denote by $w(n,h)$ the mean range in samples of n items classified in groups of breadth h randomly placed, we have by definition

$$w(n,h) = \sum_{m=0}^{\infty} \phi(n,h,m)\, mh = \sum_{m=0}^{\infty} \{P(n,h,m) - P(n,h,m-1)\}\, mh. \tag{26}$$

We now introduce central differences of the function $F_n(W)$ and use the notation

$$\Delta'_{m+\frac{1}{2}} = F_n(\overline{m+1}h) - F_n(mh), \quad \Delta''_m = F_n(\overline{m+1}h) - 2F_n(mh) + F_n(\overline{m-1}h), \tag{27}$$

so that we obtain from (25), (26) and (27)

$$w(n,h) = \sum_{m=1}^{\infty} \Delta''_m\, m.$$

We may now write

$$\begin{aligned} w(n,h) &= \lim_{M\to\infty} \sum_{m=1}^{M} \Delta''_m\, m \\ &= \lim_{M\to\infty} \left\{ M\Delta'_{M+\frac{1}{2}} - \sum_{m=0}^{M-1} \Delta'_{m+\frac{1}{2}} \right\} \\ &= \lim_{M\to\infty} \{ M\Delta'_{M+\frac{1}{2}} - F_n(Mh) \}. \end{aligned} \tag{28}$$

On the other hand, we have for the mean true range ($\overline{w}_n$ say)

$$\overline{w}_n = \int_0^{\infty} w f_n(w)\, dw,$$

where $f_n(w) = \dfrac{d}{dw} P_n(w)$ is the distribution function of the true range. We, therefore, have

$$\begin{aligned} \overline{w}_n &= \lim_{M\to\infty} \int_0^{Mh} w f_n(w)\, dw \\ &= \lim_{M\to\infty} \left\{ MhP_n(Mh) - \int_0^{Mh} P_n(w)\, dw \right\}. \end{aligned} \tag{29}$$

On taking the difference of (28) and (29) we see that

$$w(n,h) - \overline{w}_n = \lim_{M\to\infty} Mh \left\{ \frac{1}{h} \Delta'_{M+\frac{1}{2}} - P_n(Mh) \right\}.$$

Now $$\frac{1}{h} \Delta'_{M+\frac{1}{2}} = \frac{1}{h} \int_{Mh}^{(M+1)h} P_n(w)\, dw = P_n(Mh) + f_n(w^*)\, h^*,$$

where $$Mh < w^* < (M+1)\, h \quad \text{and} \quad h^* \leqslant h,$$

so that $$|\, w(n,h) - \overline{w}_n \,| \leqslant \lim_{M\to\infty} f_n(w^*)\, Mh^2.$$

From this relation it is obvious that $w(n,h)=\overline{w}_n$, since $f_n(w)$ has contact of at least second order as $w\to\infty$.*

We have proved, therefore, that if grouping is random the expectation† of the mean range $w(n,h)$ (mean grouped range) is identical with the true mean range $\overline{w}_n$ so that no bias is introduced through random grouping, no matter how large the grouping interval h. However, if we wish to use $\xi_n-\xi_1$ as an estimate of w_n this estimate, although unbiased in the sense defined, becomes less and less reliable as h increases. This is borne out by its random sampling distribution or its probability integral $P(n,h,m)$ given by (25). For normal samples it is an easy matter to tabulate $P(n,h,m)$ from the table of $P_n(W)$ (pp. 302–7) and thus to follow up the numerical increase of its standard deviation as h increases. However, to cover the case of a general parental distribution $f(x)$, we shall derive an analytical formula for the variance of $\xi_n-\xi_1$ from which approximate numerical results are easily obtained.

In order to obtain this formula we consider the second moment of $\xi_n-\xi_1$ which we may denote by $\mu_2(n,h)$. We have by definition and from equations (25), (26) and (27)

$$\mu_2(n,h)=\lim_{M\to\infty}\sum_{m=0}^{M} m^2h^2\frac{1}{h}\Delta''_m$$

$$=\lim_{M\to\infty}\sum_{m=0}^{M} hm(m+\tfrac{1}{2})\Delta''_m-\tfrac{1}{2}h\overline{w}_n. \tag{30}$$

Now we may write

$$\sum_{m=0}^{M} m(m+\tfrac{1}{2})\Delta''_m=\sum_{m=0}^{M} m\{(m+\tfrac{1}{2})\Delta'_{m+\frac{1}{2}}-(m-\tfrac{1}{2})\Delta'_{m-\frac{1}{2}}\}-\sum_{m=0}^{M} m\Delta'_{m-\frac{1}{2}}$$

$$=M(M+\tfrac{1}{2})\Delta'_{M+\frac{1}{2}}-\sum_{m=0}^{M-1}(m+\tfrac{1}{2})\Delta'_{m+\frac{1}{2}}-MF_n(Mh)+\sum_{m=0}^{M-1}F_n(mh)$$

$$=M(M+\tfrac{1}{2})\Delta'_{M+\frac{1}{2}}-(2M-\tfrac{1}{2})F_n(Mh)+2\sum_{m=0}^{M-1}F_n(mh). \tag{31}$$

Using equation (31) we obtain for the second moment (30)

$$\mu_2(n,h)=\lim_{M\to\infty}\left\{(Mh)^2\left(1+\frac{1}{2M}\right)-2MhF_n(Mh)\left(1-\frac{1}{4M}\right)+2h\sum_{m=0}^{M-1}F_n(mh)-\frac{h}{2}\overline{w}_n\right\}. \tag{32}$$

This formula enables us to compare $\mu_2(n,h)$ with $\mu_2(n)$, the second moment of the distribution of the true range. We have by definition

$$\mu_2(n)=\lim_{M\to\infty}\int_0^{Mh} f_n(w)\,w^2\,dw,$$

* It can be proved that the order of contact of $f_n(w)$ is the same as that of the parental distribution $f(x)$.

† If repeated samples were drawn from the same population and the same grouping system used in each case, the mean grouped range would be biased by an unknown amount. But in repeated experience with different populations the expectation of this bias is zero.

which we transform by two partial integrations into the equation

$$\mu_2(n) = \lim_{M\to\infty}\left\{(Mh)^2 P_n(Mh) - 2(Mh)\,F_n(Mh) + 2\int_0^{Mh} F_n(w)\,dw\right\}. \tag{33}$$

Taking the difference of (32) and (33) we obtain

$$\mu_2(n,h)-\mu_2(n) = \lim_{M\to\infty}\left\{\frac{Mh}{2}h - \frac{hF_n(Mh)}{2} - \frac{h}{2}\bar{w}_n\right\}$$
$$+\lim_{M\to\infty}\left\{2h\sum_{m=0}^{M-1}F_n(mh) + hF_n(Mh) - 2\int_0^{Mh}F_n(w)\,dw\right\}. \tag{34}$$

Since the expected mean value of the grouped range is the same as for the true range, the difference in second moments about zero equals the difference in variances. The first term on the right-hand side of equation (34) is obviously 0 (see equation (29)), whilst the second term is best evaluated with the help of Gregory's formula for numerical integration (see e.g. L. J. Comrie, 1936, p. 809). Using this formula we can express the difference between the integral and the finite sum in equation (34) in terms of the differences of the integrand $F_n(w)$ at the two ends of the range of integration. We obtain

$$\mu_2(n,h)-\mu_2(n) = \lim_{M\to\infty}\left\{\frac{2h}{12}\{\Delta'_{M-\frac{1}{2}}-\Delta'_{\frac{1}{2}}\}+\frac{2h}{24}\{\Delta''_{M-1}+\Delta''_1\}+\dots\right\}$$
$$= \tfrac{1}{6}h\,(h-\Delta'_{\frac{1}{2}}) + \frac{1}{12}h\Delta''_1 + \dots, \tag{35}$$

provided the Gregory expansion is convergent.* $\Delta'_{\frac{1}{2}}, \Delta''_1, \Delta'''_{\frac{3}{2}}, \dots$ are advancing differences of the function $F_n(w)$ at $w = 0$.

Equation (35) yields the desired formula for the second moment of $\xi_n - \xi_1$. For most parental distributions the resulting probability integral of the range will be practically 0 for a certain range in the neighbourhood of the origin (see for instance the behaviour of $P_n(W)$ from a normal parent given in the table on pp. 302–7). For such parents and for moderate h we have

$$\Delta'_{\frac{1}{2}} \cong \Delta''_1 \cong \Delta'''_{\frac{3}{2}} \cong \dots \cong 0,$$

so that
$$\mu_2(n,h)-\mu_2(n) \cong \tfrac{1}{6}h^2. \tag{36}$$

For small or moderate values of h, therefore, the increase in variance of the grouped range is given approximately (and for most parents to a high degree of accuracy) by $\frac{1}{6}h^2$. This increase is double the amount given by the well-known Sheppards correction of $\frac{1}{12}h^2$. Indeed, had we grouped a *sample of true ranges* w in fixed categories of breadth h, the resulting second moment of the grouped distribution of w would have an expectation which is $\frac{1}{12}h^2$ in excess of the second moment of the true range. With *random grouping* of the *original sample* (as it

* This condition is as a rule fulfilled for values of h which do not exceed the standard deviation of the parental distribution $f(x)$.

has been defined above in accordance with common practice of grouping) an additional uncertainty is introduced by using a new, randomly selected, set of group intervals each time a new grouped range is determined. This additional uncertainty has been proved roughly to double the excess of the variance and the result is an increase of $\frac{1}{6}h^2$ over the variance of the true range.

I wish to acknowledge with gratitude the helpful suggestions and criticisms made by Drs J. Wishart and J. O. Irwin and by Professor E. S. Pearson at various stages of this investigation.

REFERENCES

Comrie, L. J. (1936). Interpolation and Allied Tables. H.M.S.O.
McKay, A. T. (1935). *Biometrika*, **27**, 466–71.
McKay, A. T. & Pearson, E. S. (1933). *Biometrika*, **25**, 415–20.
Pearson, E. S. (1926). *Biometrika*, **18**, 173–94.
—— (1932). *Biometrika*, **24**, 404–17.
Pearson, E. S. & Haines, J. (1935). *J.R. Statist. Soc.* Suppl. **2**, 83–98.
Tildesley, M. L. (1940). *Man*, **40**, 180–9.
Tippett, L. H. C. (1925). *Biometrika*, **17**, 364–87.

MISCELLANEA

(i) The Second Yearbook of Research and Statistical Methodology Books and Reviews. Edited by OSCAR KRISEN BUROS. The Gryphon Press, Highland Park, New Jersey, 1941. $5.

This is a second and much enlarged issue of a volume published in 1938. It contains nearly seventeen hundred review excerpts on 346 statistical and allied books (in the English language only), extracted from 283 different journals. The editor has attempted with considerable success to cover the whole field of statistical and probability theory, as well as their applications in every possible direction. He has also included reviews of a number of books on the general history of science, on scientific method and on the social relations of science on the ground that they are—or should be—of general interest to scientific workers in every special field. Included in this category are books such as J. D. Bernal's *The Social Function of Science*, J. G. Crowther's *The Social Relations of Science* and J. B. S. Haldane's *The Marxist Philosophy and the Sciences*.

This large volume of several hundred pages has been admirably produced and arranged. It is intended to publish a fresh volume every two years containing reviews that have appeared in the interval. The Preface sets out a variety of reasons which, in the Editor's opinion, justify the present venture and even its enlargement in the future if sufficient support is forthcoming; at the same time frank expressions of opinions are asked for from readers and reviewers.

The objectives of the *Yearbook* as set out may be classed under four general heads:

(1) To help students, teachers and librarians to select text-books with greater discrimination and to point out to them the weak and strong points of particular books.

(2) To indicate the width of the subject of statistics and the many fields in which it is applied.

(3) To make students and teachers aware of the inadequacy of much that is now presented in text books and classes; to discourage the publication of books written by persons ignorant of the latest developments in their subject.

(4) To improve the quality of reviews by encouraging editors and reviewers alike to take their responsibilities more seriously.

With the last three objectives it is hardly possible to quarrel, and it is likely that the wide circulation of this volume would provide one of the most direct methods of attaining these ends. The first objective is, however, presumably the most important, and there are bound to be differences of opinion on the probable success of the book in this direction. In the ordinary event the teacher will no doubt be made aware of new books in the field with which he is concerned by reading the notices in one or two journals specially devoted to his subject. Having obtained a suggestion of a likely book he must surely get hold of it and determine by reading it himself whether it is suitable and abreast of the latest developments. If he is not competent to do this, but must base his decision on the advice of 6–10 reviewers, it seems doubtful whether he should be teaching the subject at all.

After reading through the reviews on some dozen books contained in the present *Yearbook*, I am inclined to the following conclusions. Regarding books of outstanding but perhaps rather controversial character, as those of Harold Jeffreys and Richard von Mises, the reader will certainly gain a useful impression from the collected reviews. This is partly because in such cases the standing of the reviewers is high and their reviews interesting and fairly written, even if critical. But in the case of the more elementary text book, the position is rather different. Quite often the opinions expressed are diametrically opposite. In cases where I knew nothing of the book or its author I found myself inevitably forced to form an opinion from my own personal knowledge of the experience, the special interests

and even the character of the reviewer. Such inside information will generally not be possessed by the College instructor and certainly not by the student. We cannot, I think, escape the conclusion that the teacher who has to select a text book for his students must be competent to decide on its merits himself and, if he is not, he may be only confused by the varied opinions contained in the *Yearbook*.

Three future directions in which the volume might be enlarged are contemplated:

(1) The inclusion of reviews of foreign language (i.e. not English) books.

(2) The addition of a section devoted to non-critical abstracts of periodical literature on research and statistical methods.

(3) The publication of original criticisms by one or more persons (according to the importance and controversial nature) of articles and papers in the periodical literature.

The first addition is clearly desirable; the publication of translations of reviews in foreign journals of our own American and British books would probably be useful too; it would help us to see ourselves as others see us. With regard to the second and third proposals, the great difficulty is of course to secure the services of sufficiently competent abstractors or critics for so large an undertaking. If, quoting the Editor, the statistical student and teacher are to be kept 'abreast of modern developments in statistical theory'; to be warned 'to ignore much of the literature which either presents nothing new or presents inefficient or incorrect methods of statistical analysis'; to be told what are 'sloppy, valueless, and erroneous articles' and what are 'well-written, significant contributions', it is clear that a very great responsibility will lie on the Editor of the *Yearbook* and his collaborators. As Prof. Buros indicates, the organizing and editing of such a comprehensive service would need the support of a foundation interested in fostering the advancement of research. Indeed, to avoid duplication the organization must be built up on an international basis, possibly in collaboration with such bodies as the American Statistical Association and the Royal Statistical Society between whose representatives some discussion of a similar project took place a few years ago.

E. S. P.

BIOMETRIKA, Vol. XXXII, Parts III & IV

CONTENTS

A volume of *Biometrika* containing about 400 pages, with plates and tables, is normally issued annually. Owing to war conditions, however, delay in issue is inevitable.

Papers for publication should either be sent to

PROFESSOR E. S. PEARSON, Department of Statistics, University College, London, W.C.1,

or if more convenient may be submitted through a member of the Editorial Committee, viz.

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 25 copies of their papers free. Joint authors 15 copies each. The price of additional copies which should be ordered when the author's proof is returned will depend upon the number of pages, plates, etc.

The subscription price, *payable in advance,* is Inland 45*s. net* per volume and Abroad 54*s. net* (including packing and postage). Owing to the scarcity of early volumes, the following rates must now be charged for complete sets. Vols. 1—XXXII, including XX[B]: £131 in buckram, £120 in wrappers, not including postage. Recent volumes may still be obtained at the wrapper price; this is 64*s.* inland, including postage. Standard buckram cases with Darwin block, price 3*s.* 6*d.* + the postage per volume. Index to Vols. I to V, 2*s. net.* Index to Vols. I to XV, 5*s. net.* Cheques must be made payable to *Biometrika* and sent to The Secretary, *Biometrika Office*, Department of Statistics, University College, London, W.C.1, to whom all orders for series, single copies and offprints should be addressed. All cheques must be *properly stamped* and should be crossed "*a/c Biometrika Trust*". No foreign cheques can be accepted unless they are drawn in sterling and payable at a London agency.